DNA replication-mediated error correction of ectopic CENP-A deposition maintains centromere identity

Yael Nechemia-Arbely[1*], Karen H. Miga[2*], Ofer Shoshani[1], Aaron Aslanian[3], Moira A. McMahon[1], Ah Young Lee[1], Daniele Fachinetti[1, 4], John R. Yates III[3], Bing Ren[1], and Don W. Cleveland[1#]

* These authors contributed equally

[1]Ludwig Institute for Cancer Research and Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA 92093 USA

[2]Center for Biomolecular Science & Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA

[3]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA

[4]Present address: Institut Curie, PSL Research University, CNRS, UMR 144, 26 rue d'Ulm, F-75005, Paris, France

[#]Corresponding author

Abstract 1 (100 words)

Chromatin assembled with the histone H3 variant CENP-A is the heritable epigenetic determinant of human centromere identity. Using genome-wide mapping and reference models for 23 human centromeres, CENP-A is shown in early G1 to be assembled into nucleosomes within megabase, repetitive α-satellite DNAs at each centromere and onto 11,390 transcriptionally active sites on the chromosome arms. Here we identify that DNA replication acts as an error correction mechanism to sustain centromere identity through the removal of the sites of CENP-A loading on the chromosome arms, while maintaining centromere-bound CENP-A with the same DNA sequence preferences as in its initial loading.

Introduction

Correct chromosome segregation during mitosis is crucial to ensure each daughter cell will receive a complete set of chromosomes. This process relies on a unique chromatin domain known as the centromere. Human centromeres are located on megabase long [1] chromosomal regions and are comprised of tandemly repeated arrays of a ~171 bp element, termed α-satellite DNA [2-4]. CENP-A is a histone H3 variant [5,6] that replaces histone H3 in ~3% of α-satellite DNA repeats [7,8], is flanked by pericentric heterochromatin containing H3K9me2/3 [9], and apparently spans on 1/3rd to one half of α-satellite arrays in the centromeres of chromosomes X and Y [10]. Despite the correlation between centromere location and the presence of α-satellite DNA repeats, α-satellite DNA sequences are neither sufficient nor essential for centromere identity [2,11,12]. This has been demonstrated by several measures including identification of multiple examples [13] of acquisition of a new centromere (referred to as a neocentromere) at a new location coupled with inactivation of the original centromere on the same chromosome. Indeed, while α-satellite arrays incorporated into human artificial chromosomes (HACs) can nucleate active centromeres [14-21], they do so at low (5-8%) efficiencies.

All of this has led to a consensus view that mammalian centromeres are defined by an epigenetic mark [2]. Use of gene replacement in human cells and fission yeast has identified the mark to be CENP-A-containing chromatin [22], which maintains and propagates centromere function indefinitely by recruiting CENP-C and the constitutive centromere associated complex (CCAN) [23-26]. We [8] and others [27] have shown that the overwhelming majority of human CENP-A chromatin particles contain two molecules of CENP-A at all cell cycle points, with CENP-A chromatin bound at authentic centromeres protecting 133 bp of centromeric α-satellite-containing DNA from nuclease digestion [8,28] before and after DNA replication [8]. This evidence

is consistent with an octameric nucleosome with DNA unwinding at all cell cycle points, and with no evidence for oscillation between hemisomes and octasomes, and with heterotypic CENP-A/histone H3-containing nucleosomes comprising at most 2% of CENP-A-containing chromatin [8].

During DNA replication, initially bound CENP-A is quantitatively redistributed to each daughter centromere [29], while incorporation of new molecules of CENP-A into chromatin occurs only for a short period after exit from mitosis [29-32] when its loading chaperone HJURP [33,34] is active [35]. This temporal separation of new CENP-A chromatin assembly at mitotic exit from centromeric DNA replication raises the important question of how is the epigenetic mark that determines centromere identity maintained across the cell cycle when it is expected to be dislodged by DNA replication and diluted at each centromere as no new CENP-A is assembled until the next G1 [29]. Moreover, endogenous CENP-A comprises only ~0.1% of the total histone H3 variants. Recognizing that a proportion of CENP-A is assembled at the centromeres with the remainder loaded onto sites on the chromosome arms [7,8,36], long-term maintenance of centromere identity and function requires limiting accumulation of non-centromeric CENP-A. Indeed, artificially increasing CENP-A expression by several fold in human cells [36-39] or flies [40] or the CENP-A homolog (Cse4) in yeast [41,42] increases ectopic deposition at non-centromeric sites, accompanied by chromosome segregation aberrations.

Using centromere reference models for each of the centromeres of the 22 human autosomes and the X chromosome, we show that after DNA replication centromere-bound CENP-A is reassembled into nucleosomes onto α-satellite DNA sequences with sequence preferences that are indistinguishable from those bound in its initial HJURP-dependent loading at mitotic exit and that this re-loading is independent of CENP-A expression level. Furthermore, we identify that a

DNA synthesis-mediated error correction mechanism acts in S phase to remove ectopically loaded CENP-A found within transcriptionally active chromatin outside of the centromeres while retaining centromere-bound CENP-A, resulting in maintenance of epigenetically defined centromere identity.

Results

CENP-A binding at 23 human centromere reference models

To identify the sequences bound by CENP-A across each human centromere, chromatin was isolated from synchronized HeLa cells expressing either i) CENP-A^{LAP}, a CENP-A variant carboxy-terminally fused to a localization [EYFP] and affinity [His] purification tag [43] at one endogenous CENP-A allele (Fig. S1a, Fig. 1a) or ii) stably expressing an elevated level (4.5 times the level of CENP-A in parental cells) of CENP-A^{TAP}, a CENP-A fusion with carboxy-terminal tandem affinity purification (S protein and protein A) tags separated by a TEV protease cleavage site (Fig. S1b, Fig. 1a). Centromeric localization of both CENP-A variants was confirmed using immunofluorescence (Fig. S1c, d), each of which has previously been shown to support long-term centromere identity and mediate high fidelity chromosome segregation in the absence of wild type CENP-A [7,8] .

Chromatin was isolated from cells synchronized to be in G1 or G2 (Fig. 1a). Synchronization efficiency was >80% as determined by fluorescent activated cell sorting (FACS) analysis of DNA content (Fig. S1e). In parallel, chromatin was also isolated from randomly cycling cells stably expressing TAP tagged histone H3.1 (H3.1^{TAP} – Fig. S1b, Fig. 1a) [23]. Chromatin from each line at G1 and G2 cell cycle phases was digested with micrococcal nuclease to generate mono-nucleosomes, producing the expected 147 bp of protected DNA length for bulk nucleosomes assembled with histone H3 (Fig. 1a, c - upper panel). CENP-A^{LAP}, CENP-A^{TAP} or

H3.1^{TAP} containing complexes were then affinity purified from the mono-nucleosome-containing pool using anti-GFP or rabbit-IgG antibodies coupled to magnetic beads. PreScission or TEV protease cleavage was then used to elute His or S-protein tagged CENP-A or S-tagged H3.1 chromatin under mild conditions that maintain initially assembled chromatin (Fig. 1a). α-satellite DNA sequences were enriched 30-35 fold in DNA isolated from CENP-A^{TAP} or CENP-A$^{+/LAP}$ cells (Fig. 1b), the expected enrichment since α-satellite DNA comprises ~3% of the genome [8,21]. While micro-capillary electrophoresis of bulk input chromatin produced the expected 147 bp of protected DNA length for nucleosomes assembled with histone H3 (Fig. 1c - upper panel), isolated CENP-A^{LAP} chromatin expressed at endogenous CENP-A levels produced DNA lengths centered on 133 bp, both before and after DNA replication (Fig. 1c, lower panel), a distribution indistinguishable from that previously reported for octameric CENP-A-containing nucleosomes assembled in vitro and in which DNA unwinding at entry and exit has been demonstrated [8,44].

Libraries of affinity purified CENP-A^{LAP}, CENP-A^{TAP}, and H3.1^{TAP}-bound DNAs were prepared, sequenced (using paired-end 100 bp sequencing), and mapped (Fig. 1a, d and Table S1) onto the published reference model for the centromere of the X chromosome [45] and unpublished reference models for the centromeres of the 22 human autosomes [46,47]. These centromere models include the observed variation in α-satellite Higher Order Repeat (HOR) array sequences contained in the HuRef genome [48]. The highly repeated sequences preclude distinguishing between centromeric and pericentromeric sequences and the order of repeats in the models is arbitrarily assigned and portions of the centromeres of the acrocentric chromosomes 13, 14, 21 and 22, as well as portions of centromeres of chromosomes 1, 5 and 19, contain nearly identical arrays that cannot be distinguished.

Sequences in each reference centromere associated with CENP-A binding were identified (Fig. 1d,e; Fig. S2) using algorithm-based scripts [SICER and MACS [49,50]]. Mapping of CENP-A^{LAP} expressed at endogenous levels across the centromeric regions of all 23 reference centromeres (see Fig. 1e for chromosome 18 and Figure S2 for the other 22) revealed the profile of CENP-A binding. Mapping to the sequences in the reference centromeres was highly reproducible (e.g., compare duplicates in Figs. 1e) and largely unaffected by increasing CENP-A levels by 4.5-fold (e.g., compare sequences bound in CENP-A^{TAP} cells with those in CENP-$A^{+/LAP}$ cells in Figs. 1e and 2a, b).

Analysis of CENP-A-bound DNAs aligned to α-satellite sequences in CENP-$A^{+/LAP}$ G1 cells revealed that our CENP-A ChIP-seq approach resulted in varying levels of array enrichment, from ~10.5-fold enrichment for array D3Z1 in cen3 to ~213-fold for array GJ211930.1 in cen10, with most active arrays showing enrichment level of 20-40-fold above background (input) level (Table S2; columns 13, 14). For the 6 of the 17 centromeres which contain more than one α-satellite array within them, CENP-A was enriched only in one, indicating that only one array was active (see centromere reference models 3, 7, 12, 15, 16 and 19; Table S2: columns 13, 14). Multiple α-satellite arrays in 11 centromeres (for example, see centromere reference models 10,11,13, 14, and 17; Table S2: columns 13, 14) showed enrichment of CENP-A binding in two or more arrays. These may represent functional epialleles for CENP-A binding, (i.e., in which CENP-A binds to a different array in each homologue), as was shown previously for cen17 in two diploid cell human lines [20]. Increased levels of CENP-A expression (in CENP-A^{TAP} versus CENP-$A^{+/LAP}$ cells) did not increase the number of binding peaks in the reference centromeres (Fig. 1d, e), but did increase occupancy within the cell population at some divergent monomeric α-satellite repeats (Fig. S1f) or within HORs (Fig. S1g), with both examples

153 occurring in regions with few 17 bp CENP-B boxes (Fig. S1f, g – compare sites bound at
154 endogenous or increased CENP-A levels) that are direct binding sites for CENP-B [51], the only
155 known sequence-specific human centromere binding protein.

156 **CENP-A nucleosomes are retained at centromeric loading sites after DNA replication**

157 Next, we examined how centromeric CENP-A binding was affected by DNA replication. Despite
158 the known redistribution of initially centromere-bound CENP-A onto each of the new daughter
159 centromeres without addition of new CENP-A [29], comparison of the sequences bound by CENP-
160 A in G1 with those bound in G2 revealed the remarkable feature that for all 23 centromeres, at
161 both normal (CENP-$A^{+/LAP}$) and elevated (CENP-A^{TAP}) levels, CENP-A was bound to
162 indistinguishable α-satellite sequences before and after DNA replication (shown for the
163 reference centromere of chromosome 18 in Fig. 1e and for the other chromosomes in Fig. S2).
164 Indeed, most (87%) of α–satellite sequences with CENP-A binding peaks in chromatin
165 immunopurified from G1 CENP-A^{TAP} cells remained at G2 (Fig. S1h, top). Similarly high overlap
166 (89%) was identified between repeats (called by SICER [50]) bound by CENP-A in G1 or G2 when
167 CENP-A was expressed at endogenous levels (in CENP-$A^{+/LAP}$ cells) (Fig. S1h, bottom).

168 To address whether CENP-A was precisely retained at the same centromeric α–satellite loading
169 sites after DNA replication, we filtered out multi-mapping reads, leaving only reads that map to
170 sites that are single copy in the HuRef genome, and therefore are mapped uniquely. Analysis
171 of these single copy variants revealed that CENP-A was quantitatively retained after DNA
172 replication at single copy sites found within HORs (Fig. 2a-f). Indeed, almost all (93 of 96) unique
173 CENP-A binding sites mapped within the 23 centromeres of CENP-A^{TAP} cells in G1 remained
174 undiminished at G2 (Fig. 2g, left), with the remaining 3 peaks only slightly diminished.

175 **Sites of CENP-A assembly onto chromosome arms in early G1 are removed by G2**

Genome-wide mapping of CENP-A-bound DNAs revealed that, in addition to the striking enrichment at centromeric α-satellites, CENP-A was preferentially and highly reproducibly incorporated into unique sequence, non-α-satellite sites on the arms of all 23 chromosomes (Figs. 3a, b and 4). Sites enriched for CENP-A binding were essentially identical in DNAs from randomly cycling cells or G1 cells (Fig. 3a, b, mapped sequence reads in color and binding sites underneath in black). A 4.5-fold increase in CENP-A levels in CENP-A^{TAP} cells drove correspondingly increased CENP-A sites of incorporation on the arms (from 11,390 sites at normal CENP-A levels to 40,279 sites when CENP-A was elevated 4.5-fold - Fig. 3a-b, d), but did not increase the binding peaks within the centromeric HORs (Fig. 1d) or the number of unique single copy sites within centromeric HORs (Fig. 2g, right).

Remarkably, for all 23 human chromosomes and for CENP-A accumulated to endogenous (CENP-A^{LAP}) or increased (CENP-A^{TAP}) expression levels, passage from G1 to G2 almost eliminated enrichment of CENP-A binding to specific sites on the chromosome arms, while leaving α-satellite bound sequences unaffected (Figs. 1d, 3, 4). Loss by G2 of CENP-A binding in G1 at specific arm sites was highly reproducible, as demonstrated by experimental replicas (Fig. 3b). On a genome-wide scale, scoring peak binding sites with thresholds of ≥ 5-fold, 10-fold or 100-fold of CENP-A binding over background, at least 90% of sites bound on chromosome arms in G1 in CENP-A^{TAP} cells were removed by early G2 (Fig. 3e) and all of those still identified in G2 were substantially reduced in peak height.

Taken together, whether at endogenous or increased CENP-A expression level, CENP-A is loaded in G1 not only at centromeric α-satellite DNAs but also at preferential sites on the chromosome arms, but passage across S phase removes or sharply diminishes these unique sequence sites of enhanced CENP-A binding on the arms, while CENP-A binding to

centromeric sequences is retained.

CENP-A loading sites on chromosome arms are not seeding hotspots for neocentromere formation

We next determined whether sites of preferential CENP-A loading onto the chromosome arms corresponded to sites that have become active "neocentromere" locations [52] for the three human neocentromeres for which prior work has defined their chromosomal locations [28]. The first of these, named PDNC4, has a neocentromere on chromosome 4 [53] that spans 300 kb (from 88.2 to 88.5 Mb). No peak binding sites of CENP-A relative to neighboring regions were found in this chromosomal region even with elevated CENP-A expression in CENP-A^{TAP} cells (Fig. 5a). Similar examination of two additional neocentromere-containing locations/positions [line MS4221 that harbors a 400 kb neocentromere at position 86.5 to 86.9 Mb on chromosome 8 [28,54] and line IMS13q with a neocentromere on chromosome 13 [55] that spans 100 kb (from 97.7 to 97.8 Mb)] again revealed no elevated incorporation of CENP-A^{TAP} within the DNA sequences corresponding to the neocentromere domains (Fig. S3a, b).

CENP-A is ectopically loaded at early G1 into open and active chromatin

We examined the nature of the sites on the chromosome arms into which CENP-A was assembled in G1. A 2-fold enrichment (compared to levels expected by chance) of CENP-A^{TAP} bound to unique arms sites during G1 was found at promoters, enhancers or promoters of expressed genes and a 2.5-fold enrichment at sites bound by the transcriptional repressor CTCF (Fig. S3d), a similar trend to what has been observed previously for increased expression of CENP-A [36]. Importantly, more than 80% of CENP-A^{TAP} binding sites on chromosome arms with peak heights ≥ 5-fold over background overlapped with DNase I hypersensitive sites identified by comparison with ENCODE DNase I hypersensitive datasets (with minimum overlap

of 100 bases) that denote accessible chromatin zones and which are functionally related to transcriptional activity (Fig. 5b, c). A ~5.5-fold enrichment of CENP-A^{TAP} was found at these sites (Fig. 5d). CENP-A assembled into chromatin when expressed at endogenous levels was also found to be enriched 3-fold at DNase I hypersensitive sites (Fig. 5e) and promoters (Fig. S3f). Conversely, both CENP-A^{TAP} and CENP-A^{LAP} were not enriched at H3k27me3 peak sites that are tightly associated with inactive gene promoters (Fig. S3c, e).

Ectopic CENP-A is removed contemporaneously with replication fork progression, while centromeric CENP-A is retained

We next tested whether removal by G2 of CENP-A assembled into nucleosomes at unique sites on the chromosome arms is mediated by the direct action of the DNA replication machinery. CENP-A^{TAP} was affinity purified from mid S phase cells and CENP-A-bound DNAs were sequenced (Fig. 6a; Fig. S4a). In parallel, we pulse-labeled newly synthesized DNA in our synchronized cells by addition of bromodeoxyuridine (BrdU) for 1 hour at early (S0-S1), mid (S3-S4), and late S phase (S6-S7) (Fig. 6a; Fig. S4a). Genomic DNA from each time point was sonicated (Fig. S4b) and immunoprecipitated with a BrdU antibody (Fig. 6a). Eluted DNA was then sequenced and mapped to the genome [an approach known as Repli-seq [56]], yielding regions of early, mid, and late replicating chromatin (an example from a region of chromosome 20 arm is shown in Fig. 6b). The early replication timing was validated (using qPCR - Fig. S4c) for two genes (MRGPRE and MMP15) previously reported to be early replicating [ref [57] and ENCODE Repli-seq]. Similarly, a gene and a centromeric region (HBE1 and Sat2) previously reported to be late replicating [ref [58] and ENCODE Repli-seq] were confirmed in our cells to be replicated late (Fig. S4d).

CENP-A immunoprecipitation from micrococcal nuclease digested chromatin isolated from

246 early and mid S phase cells resulted in levels of α-satellite DNA enrichment (Fig. S4e) similar
247 to those achieved at G1 phase (Fig. 1b). Furthermore, nucleosomal CENP-A chromatin
248 produced by micrococcal nuclease digestion protected 133 bp of DNA at early and mid S phase
249 (Fig. S4f) just as it did in G1 and G2 [Fig. 1c; see also ref [8]], with no evidence for a structural
250 change from hemisomes to nucleosomes and back to hemisomes during S phase as previously
251 claimed [59]. Mapping of CENP-A binding sites within the chromosome arms, combined with
252 Repli-seq analysis, revealed that 91% of ectopic G1 CENP-A binding sites were found in early-
253 or mid-S replicating regions (Fig. 6b, c). While centromeres of chromosomes 1, 3, 10, 17, 18
254 and X [60] and bulk α-satellite DNAs or a consensus pool of alphoid DNA sequences have been
255 reported to replicate at mid [61,62] or mid-to-late [63] S phase, in our cells α-satellite containing
256 DNAs in all 23 centromeres were found almost exclusively to be late replicating (Fig. 6d).

257 Remarkably, throughout S phase, centromere bound CENP-A found in G1 was completely
258 retained across each reference centromere with the same sequence binding preferences
259 (shown for centromere 18 in Fig. 6e and Fig. S4g). Retention of CENP-A binding during DNA
260 replication was observed also at the unique sequence binding sites within HORs (Fig. 6f).
261 Indeed, all 96 CENP-A^{TAP} G1 peaks at single copy variants within α-satellite HORs remained
262 bound by CENP-A at mid S phase (Fig. 6g). In contrast, early replicating ectopic CENP-A
263 binding sites were nearly quantitatively removed during or quickly after their replication and
264 were no longer visible at mid-S phase (Fig. 6h, i). Similarly, ectopic CENP-A binding sites that
265 were in mid-S replicating regions remained at mid-S but were removed quickly after that and
266 were absent by late S/G2 (Fig. 6j, k). Ten percent of ectopic CENP-A G1 peaks were in late-S
267 replicating regions (Fig. 6c). Here again, almost all (85%) of these were removed by G2 (Fig.
268 6l, m), while late replicating centromeric CENP-A peaks were retained (Fig 6d-g). These results

demonstrate that ectopic, but not centromeric, CENP-A binding sites are removed as DNA replication progresses. Moreover, that contemporaneously late replicating centromeric bound CENP-As, including the unique binding sites within the 23 centromeres, were retained following DNA replication while ectopic sites were removed eliminates the possibility that retention could be a consequence of a general alteration in late S phase in the activity of one or more DNA replication components that could potentially act to facilitate CENP-A reloading.

Centromeric CENP-A is continuously bound by the CCAN complex during centromeric DNA replication

To comprehensively determine the components which associate with CENP-A during replication in late S, we used mass spectrometry following affinity purification of CENP-A nucleosomes (Fig. S4h, left panel). A structural link that normally bridges multiple centromeric CENP-A nucleosomes and nucleates full kinetochore assembly before mitotic entry [64-70] is the 16-subunit constitutive centromere associated network (CCAN). This complex is anchored to CENP-A primarily through CENP-C [68,71-75] and sustained by CENP-B binding to α-satellites [76].

Remarkably, mass spectrometry identified that all 16 CCAN components [23,25] remained associated with mono-nucleosomal CENP-A chromatin affinity purified from late S/G2 (Fig.6n; Fig. S4h). Further, association of CENP-A with MCM2 (and other components of the MCM2-7 replicative helicase complex) and CAF1p150 was enhanced at late S phase (compared with its association in randomly cycling cells) (Fig. S4i). Stable association with CENP-A was also seen for HJURP, multiple chromatin remodeling factors and nuclear chaperones (Fig. S4k), histones (Fig. S4l), centromere and kinetochore components (Fig. S4m), and other DNA replication proteins (Fig. S4j).

The continuing interaction during DNA replication of CCAN proteins with CENP-A and which is maintained even on mono-nucleosomes provides strong experimental support that the CCAN network tethers CCAN-bound centromeric CENP-A at or near the centromeric DNA replication forks, thereby enabling its efficient reincorporation after replication fork passage. To test this further, the composition of CENP-A-containing nucleosomal complexes from G1 to late S/G2 was determined following affinity purification (via the TAP tag) of chromatin-bound CENP-A^{TAP} from predominantly mononucleosome pool (Fig. S4h, right panel).

We initially focused on the Chromatin Assembly Factor 1 (CAF1) complex, which is required for *de novo* chromatin assembly following DNA replication [77,78]. Its p48 subunit (also known as CAF1 subunit c, RbAp48, or RBBP4) 1) binds histone H4 [79] and 2) has been reported as a binding partner in a CENP-A pre-nucleosomal complex with HJURP and nucleophosmin (NPM1) [34]. In this latter complex it has been proposed to promote H4K5 and K12 acetylation prior to CENP-A loading [80] and maintain the deacetylated state of histones in the central core of centromeres after deposition [81]. Immunoblotting revealed that CAF1 p48 co-immunopurified with CENP-A from G1 through late S/G2 (Fig. 6o), consistent with a role for it in binding H4 and perhaps maintaining a deacetylated state.

In striking contrast, the two other CAF1 subunits, CAF1 p150 and CAF1 p60, that are essential for *de novo* chromatin assembly *in vitro* [82], remained much more strongly associated with CENP-A nucleosomes in late S/G2 than in mid-S (Fig. 6o). Additionally, MCM2, a core subunit of the DNA replicative helicase MCM2-7 complex that has an important role in recycling of old histones as the replication fork advances [83], was robustly co-purified with CENP-A only in late S phase derived chromatin, with no association detected in mid-S (Fig. 6o), when ectopic CENP-A peaks replicate. Thus, there is stable association of CENP-A with MCM2 and the CAF1

subunits necessary for chromatin reassembly after replication only in late S phase, the time when all centromeric, but only a small minority of ectopically loaded, CENP-A is replicated.

Discussion

Using reference models for 23 human centromeres, we have identified that during DNA replication CENP-A nucleosomes initially assembled onto centromeric α-satellite repeats are reassembled onto the same spectrum of α-satellite repeat sequences of each daughter centromere as was bound prior to DNA replication. Additionally, genome-wide mapping of sites of CENP-A assembly identified that when CENP-A is expressed at endogenous levels, the selectivity of the histone chaperone HJURP's loading in early G1 of new CENP-A at or near existing sites of centromeric CENP-A-containing chromatin is insufficient to prevent its loading onto >11,000 sites along the chromosome arms (Fig. 3d). We also show that the number of ectopic sites increases as CENP-A expression levels increase, as has been reported in multiple human cancers [39,63,64]. These sites of ectopic CENP-A are replicated in early and mid-S (Fig. 6b, c) and are nearly quantitatively removed as DNA replication progresses (Fig. 6h-m).

Taken together, our evidence identifies that DNA replication functions not only to duplicate centromeric DNA but also as an error correction mechanism to maintain epigenetically-defined centromere position and identity by coupling centromeric CENP-A retention with its removal from assembly sites on the chromosome arms (Fig. 6p). Indeed, our data reveal that CENP-A loaded onto unique sites (after filtering out multi-mapping reads in the α-satellite HORs in the HuRef genome) within the 23 reference centromeres, is precisely maintained at these sites during and after DNA replication, offering direct support that at least for each of these unique sites the replication machine re-loads CENP-A back onto the exact same centromeric DNA site (Fig. 2, 6f,g). Accompanying this is retention of centromeric, α-satellite DNA-bound CENP-A

337 before and after DNA replication at indistinguishable sequences throughout reference models
338 of all 23 human centromeres (Fig. S2). DNA replication produces a very different situation for
339 CENP-A initially assembled into nucleosomes on the chromosome arms. Sites of this
340 ectopically loaded CENP-A are nearly quantitatively stripped during DNA replication (Figs. 3, 4,
341 6h-m), providing strong evidence that DNA replication acts not only to duplicate both
342 centromeric and non-centromeric DNA sequences, but also to reinforce epigenetically defined
343 centromere position and identity, while precluding acquisition of CENP-A-dependent
344 centromere function at non-centromeric sites (Fig. 6p).

345 Without such correction, ectopically loaded sites would be maintained and potentially reinforced
346 cell cycle after cell cycle, ultimately recruiting CENP-C which in turn can nucleate assembly of
347 the CCAN complex [23-26]. Increasing levels of arm-associated CENP-A/CCAN would present a
348 major problem for faithful assembly and function of a single centromere/kinetochore per
349 chromosome, both by acquisition of partial centromere function and by competition with the
350 authentic centromeres for the pool of available CCAN components. Indeed, high levels of
351 CENP-A overexpression 1) leads to recruitment of detectable levels of 3 of 16 CCAN
352 components (CENP-C, CENP-N and Mis18) assembled onto the arms [36,37,84], 2) ongoing
353 chromosome segregation errors [38], and 3) has been observed in several cancers where it has
354 been associated with increased invasiveness and poor prognosis [39,85,86].

355 As to the mechanism for retention during DNA replication of centromeric but not ectopically
356 loaded CENP-A, an attractive model strongly supported by our evidence is that the local
357 reassembly of CENP-A within centromeric domains is mediated by the continuing centromeric
358 CENP-A association with CCAN complexes, which we show to be maintained on individual
359 CENP-A nucleosomes in late S when centromeres are replicated (Fig. 6n). The continued

presence of the assembled CCAN network directly bound to CENP-A during centromere DNA replication offers a plausible explanation for centromeric CENP-A retention (together with the MCM2 replicative helicase and the major CAF1 subunits required for nucleosome reassembly post-replication [77,83]). In such a model, centromere identity is preserved by an assembled CCAN network which serves during DNA replication to tether disassembled CENP-A/H4 dimers or tetramers near the sites of centromere replication, thereby enabling their local reassembly onto each of the daughter centromeres and the corresponding epigenetic inheritance of centromere identity.

Figure Legends

Figure 1. CENP-A ChIP-seq identifies CENP-A binding at reference centromeres of 23 human chromosomes. (**a**) CENP-A ChIP-sequencing experimental design. (**b**) Quantitative real-time PCR for α-satellite DNA in chromosomes 1, 3, 5, 10, 12 and 16. N=2 for CENP-A^{TAP} and 3 for CENP-A^{LAP}, from two independent replicates. Error bars, s.e.m. *P* value of 0.608 determined using two-tailed *t*-test. (**c**) MNase digestion profile showing the nucleosomal DNA length distributions of bulk input mono-nucleosomes (upper panel) and purified CENP-A^{LAP} following native ChIP at G1 and G2. (**d**) Number of CENP-A binding peaks at α-satellite DNA in CENP-A^{TAP} and CENP-$A^{+/LAP}$ cells at G1 and G2. The number represent peaks that are overlapping between the two replicates. (**e**) CENP-A ChIP-seq shows CENP-A binding peaks at the centromere of chromosome 18 for CENP-A^{LAP} and CENP-A^{TAP} before and after DNA replication. CENP-A peaks across the reference model are a result of multi-mapping and their exact linear order is not known. SICER peaks are shown in black below the raw read data. Two replicates are shown for each condition. Scale bar, 2Mb.

Figure 2. Retention of centromeric CENP-A through DNA replication. (**a, b**) CENP-A ChIP-seq raw mapping data (colored) and SICER peaks (black lines, underneath) showing sequences bound by CENP-A (at both endogenous and increased expression levels) across centromere reference model of chromosome 8, before and after DNA replication. Upper part shows mapping of all reads (including reads that are multi-mapping) onto the repetitive α–satellite DNA. Lower part shows read mapping to sites that are single copy in the HuRef genome (single-mapping), after filtering out multi-mapping reads. Centromere reference location, red. CENP-B box location, orange. Scale bar, 500kb (**a**), 10kb (**b**). (**c**) High-resolution view of read mapping to a site that is single copy in centromere reference model of chromosome

8, marked by a purple bar in (**a**). Scale bar, 200bp. (**d**) High-resolution view of read mapping to a site that is single copy in centromere reference model of chromosome 8, marked by a green bar in (**a**). Scale bar, 500bp. (**e, f**) High-resolution view of read mapping to a site that is single copy in centromere reference model of chromosome 2 (**e**) and in the centromere reference model of chromosome 10 (**f**). Scale bar, 200bp (**e**) and 500bp (**f**). (**g**) Left, overlap between G1 and G2 CENP-A single mapping binding sites at α-satellite HOR sequences. Right, peak overlap between G1 CENP-A^{TAP} (increased expression) and CENP-A^{LAP} (endogenous level) single mapping binding sites at α-satellite HOR sequences.

Figure 3. Sites of CENP-A assembly onto chromosome arms in early G1 are removed by G2. (**a**) ChIP-sequencing raw mapping data (colored) and SICER peaks (black lines, underneath) showing sequences bound by CENP-A (at both endogenous and increased expression levels) across chromosome 4 before and after DNA replication. Centromere reference location, red. CENP-B box location, orange. Read counts were scale to 30 but reaches 150 at the centromere. Scale bar, 50Mb. (**b**) ChIP-sequencing data are shown for a region within the p-arm of chromosome 4, with two replicates for each time point, for CENP-A^{TAP} (increased CENP-A expression), CENP-A^{LAP} (endogenous level), and $H3.1^{TAP}$. Scale bar, 2Mb. (**c**) High resolution nucleosomal view of CENP-A^{TAP} mapping data at G1 and G2 at a non-centromeric site of chromosome 4. Scale bar, 1kb. (**d**) Total number of non-α-satellite CENP-A binding sites for CENP-A^{TAP} and CENP-A^{LAP} at G1 and G2. The number represent peaks that are overlapping between the two replicates. (**e**) Number of non-α-satellite CENP-A SICER binding sites called at G1 or G2 at different fold thresholds (above background).

Figure 4. Ectopic CENP-A is removed following DNA replication from the arms of all 23 human chromosomes. CENP-A^{TAP} ChIP-sequencing raw mapping data at G1 (red) and G2

(blue) for all human chromosomes. Chromosome X show a spike of CENP-A enrichment not removed by G2 (marked by an asterisk).

Figure 5. Sites of deposition of CENP-A on chromosome arms are not seeding hotspots for neocentromere formation. (**a**) Read mapping data of CENP-A^{TAP} ChIP-sequencing at G1 and G2, at the chromosomal location of a known patient derived neocentromere [28] found in chromosome 4. (**b**) Greater than 80% of CENP-A SICER peaks ≥ 5-fold in randomly cycling and G1 cells overlap with DNase I hypersensitive sites taken from ENCODE project. (**c**) Example from the chromosome 4 p-arm showing overlap of at least 100 bases between SICER peaks ≥ 5-fold and HeLa S3 DNase I hypersensitive sites taken from ENCODE project. (**d, e**) CENP-A^{TAP} (d) or CENP-A^{LAP} (e) enrichement levels at DNase I hypersensitive sites. SICER peaks ≥ 5-fold supported between the two replicates were analyzed for their enrichment level at DNase I hypersensitive sites, with minimum overlap of 100 bases, compared to the level of enrichment at these sites by chance.

Figure 6. Ectopic CENP-A is removed contemporaneously with replication fork progression, while centromeric CENP-A is retained. (**a**) Schematic representation of CENP-A ChIP-seq combined with Repli-seq experiment across S phase. (**b**) Raw mapping data of CENP-A^{TAP} ChIP-seq at G1 (red) and BrdU repli-seq (grey scale) at early S (S1), mid S (S4) and late S/G2 (S7) at the q-arm of chromosome 20. SICER peaks, black lines are drawn underneath the raw mapping data. Scale bar, 5 Mb. (**c**) The percentage of ectopic G1 CENP-A^{TAP} ≥5-fold binding sites that are found in early, mid, or late S replicating regions. N=2 for from two independent replicates. Error bars, s.e.m. (**d**) Percentage of BrdU SICER peaks at α-satellite DNA found within early, mid or late S replicating regions. N=2 for from two independent replicates. Error bars, s.e.m. *P* value of 0.0011 determined using two-tailed *t*-test. (**e**) CENP-A

437 ChIP-seq raw mapping data at a part of cen18 at G1, mid S phase and G2, and BrdU repli-seq
438 at early S (S1), mid S (S4) and late S/G2 (S7). SICER peaks are denoted as black lines
439 underneath the raw mapping data. Centromere reference location, red. CENP-B boxes, orange.
440 Scale bar, 10kb. (**f**) High-resolution view of CENP-A mapping during DNA replication (mid-S) to
441 a single copy site (marked by a purple bar in Fig. 2a) in the centromere reference model of
442 chromosome 8. Data from Figure 2c for CENP-A^{TAP} G1 and G2 is included for comparison.
443 Scale bar, 200bp. (**g**) Complete overlap between CENP-A G1 and mid-S single mapping
444 binding sites (after filtering out multi-mapping reads) is found at α-satellite HORs sequences.
445 (**h, j, l**) Raw mapping data (colored) and SICER peaks (black lines, underneath) of CENP-A^{TAP}
446 ChIP-seq at G1, mid S phase and G2 and BrdU labeled repli-seq samples (grey scale) showing
447 regions going through replication at early S (S1), mid S (S4) and late S/G2 phase (S7) within
448 the p-arm of chromosome 4 (**h**), the q-arm of chromosome 10 (**j**), and the p-arm of chromosome
449 X. Early replicating CENP-A^{TAP} G1 peaks (**h, j**) are removed by mid S phase. Mid-S replicating
450 CENP-A^{TAP} G1 (**j**) peaks are removed by G2. Late replicating CENP-A^{TAP} G1 peaks (**l**) are
451 removed by G2. Scale bars, 1kb. (**i**) Early replicating CENP-A^{TAP} G1 peaks (≥5-fold over
452 background - shown in (**h**) for chromosome 4) were analyzed for their retention at mid S phase.
453 N=2 for from two independent replicates. Error bars, s.e.m. About 70% of early replicating
454 CENP-A^{TAP} G1 peaks are removed by mid S phase. (**k**) Mid-S replicating CENP-A^{TAP} G1 peaks
455 (≥5-fold over background - shown in (**j**) for chromosome 10) were analyzed for their retention
456 at G2. N=2 for from two independent replicates. Error bars, s.e.m. 90% of mid replicating CENP-
457 A peaks are removed by G2. (**m**) Late replicating CENP-A^{TAP} G1 peaks (≥5-fold over
458 background - shown in (**l**) for chromosome X) were analyzed for their retention at G2. N=2 for
459 from two independent replicates. Error bars, s.e.m. 85% of late replicating CENP-A^{TAP} G1 peaks
460 are removed by G2. (**n**) CENP-A^{TAP} was immunoprecipitated from the chromatin fraction of

randomly cycling cells or late S/G2 synchronized cells followed by mass spectrometry to identify the co-precipitated partners. All the CCAN network components were co-precipitated with CENP-A at late S/G2. (**o**) CENP-A^{TAP} was immunoprecipitated from micrococcal nuclease resistant chromatin isolated at different cell cycle phases and the immunoprecipitates were examined by immunoblotting for CAF1 complex subunits and MCM2. (**p**) Model for maintaining centromeric CENP-A while removing it from non-centromeric sites on the chromosome arms during DNA replication to ensure maintenance of centromere identity across the cell cycle.

Supplementals figure legends

Supplementary Figure S1. Identification of peaks enriched for CENP-A binding. (a) Scheme showing experimental design for tagging an endogenous CENP-A locus to produce CENP-A$^{+/LAP}$ HeLa cells. These cells were then adapted to suspension growth. (**b**) Scheme showing the experimental design for obtaining increased levels of CENP-A^{TAP} expression. CENP-A^{TAP} is expressed in these cells at 4.5-fold the level of CENP-A in the parental HeLa cells [8]. **(c, d)** Localization of endogenously tagged CENP-A^{LAP} (**c**) and CENP-A^{TAP} (**d**) determined with indirect immunofluorescence using anti-GFP antibody (**c**) or rabbit-IgG (**d**). Scale bar, 5 μm. (**e**) FACS analysis of DNA content showing the synchronization efficiency of CENP-A$^{+/LAP}$ and CENP-A^{TAP} HeLa cell lines. (**f**, **g**) Examples of centromeric regions of chromosome 7 (**f**) and 5 (**g**) showing increased occupancy of overexpressed CENP-A^{TAP} (compare CENP-A^{TAP} with CENP-A^{LAP})**.** (**h**) Overlap between G1 and G2 CENP-A binding peaks at α-satellite sequences.

Supplementary Figure S2. CENP-A ChIP-seq identifies CENP-A binding at reference centromeres of 23 human chromosomes. CENP-A^{LAP} bound DNAs at G1 and G2 were sequenced, with 2 replicates per condition, and mapped to the centromeric reference models in the hg38 assembly [46]. Shown are the raw mapping data (colored) for every human centromere (except for the centromere of chromosome 19 that shares almost all of its α-satellites arrays with α-satellites arrays of chromosomes 1 and 5) and CENP-A binding called as SICER peaks (black lines, underneath) for one replicate for each time point. Centromere reference location, red. CENP-B box, orange.

Supplementary Figure S3. Ectopic deposition of CENP-A into open and active chromatin at G1 does not function as a seeding hotspot for neocentromere formation. (**a, b**) Read

491 mapping data of CENP-A^{TAP} ChIP-sequencing at G1 (red) and G2 (blue), at the chromosomal
492 location of 2 known patient derived neocentromeres [28] found in chromosome 8 (**a**) and
493 chromosome 13 (**b**). **(c-f**) Fold enrichment of CENP-A^{TAP} chromatin in randomly cycling cells or
494 at G1 or G2 (**c, d**) and CENP-A^{LAP} chromatin (**e, f**) at G1 or G2 at different genomic locations.
495 SICER peaks ≥ 5-fold supported between two replicates were analyzed for their enrichment
496 level at different genomic locations, compared to the level of enrichment at these sites by
497 chance.

498 **Supplementary Figure S4. Centromeres are late replicating with CENP-A remaining**
499 **tethered locally by continued binding to the CCAN complex.** (**a**) FACS analysis of DNA
500 content showing the synchronization efficiency of CENP-A^{TAP} HeLa cell line across S phase.
501 (**b**) Genomic DNA of cells labeled for 1 hour with BrdU was sonicated prior to the BrdU
502 immunoprecipitation and fragments of 200-800bp were obtained. (**c**) Quantitative real-time PCR
503 for MRGPRE and MMP15 validates their early replication timing as previously reported [ref [57]
504 and ENCODE Repli-seq]. N=2, from two independent replicates. Error bars, s.e.m. (**d**)
505 Quantitative real-time PCR for HBE1 and Sat2 was used to validate their late replication timing,
506 as previously reported [ref [58] and ENCODE Repli-seq]. N=2, from two independent replicates.
507 Error bars, s.e.m. (**e**) Quantitative real-time PCR for α-satellite DNA. N=2, from two independent
508 replicates. Error bars, s.e.m. (**f**) MNase digestion profile showing the nucleosomal DNA length
509 distributions of bulk input mono-nucleosomes (upper panel) and purified CENP-A^{TAP} following
510 native ChIP at early S and mid S phase. (**g**) CENP-A ChIP-seq raw mapping data spanning the
511 whole of cen18 at G1, mid S phase and G2, and BrdU repli-seq at early S (S1), mid S (S4) and
512 late S/G2 (S7). SICER peaks are denoted as black lines underneath the raw mapping data.
513 Centromere reference location, red. CENP-B boxes, orange. Scale bar, 2Mb. (**h**) Ethidium
514 Bromide stained DNA agarose gel showing MNase digestion profile of bulk chromatin used for

515 mass spectrometry identification of proteins associating with CENP-A^{TAP} chromatin (left panel)
516 and for CENP-A^{TAP} co-immunoprecipitation experiment (right panel). (**i-m**) CENP-A^{TAP}
517 immunopurification followed by mass spectrometry identifies association with CENP-A
518 chromatin of DNA replication related proteins (**i,j**), chromatin remodeling factors and nuclear
519 chaperones (**k**), histones (**l**) and centromere and kinetochore proteins (**m**).

520

Materials and Methods

Cell lines. Adherent HeLa cells stably expressing CENP-A^{TAP} or H3.1^{TAP} by retrovirus infection [23] or endogenously tagged CENP-A$^{+/LAP}$ by infection of a rAAV harboring a LAP targeting construct containing homology arms for CENP-A [43] were adapted to suspension growth by selecting surviving cells and were maintained in DMEM medium (Gibco) containing 10% fetal bovine serum (Omega Scientific), 100U/ml penicillin, 100U/ml streptomycin and 2mM l-glutamine at 37ºC in a 5% CO_2 atmosphere with 21% oxygen. Cells were maintained and split every 4-5 days according to ATCC recommendations.

Cell synchronization. Cells were synchronized as previously described [8]. Briefly, suspension HeLa cells were treated with 2 mM thymidine in complete medium for 19 h, pelleted and washed twice in PBS, and released in complete medium containing 24 µM deoxycytidine for 9 h followed by addition of thymidine to a final concentration of 2 mM for 16 h, after which cells were released again into complete medium containing 24 µM deoxycytidine. For G2, cells were harvested 7 hours after release from the second thymidine block. For G1, thymidine was added for a third time, 7 hours after the release and cells were harvested 11 hours after that (a total of 18 hours after the release from the second thymidine block).

Chromatin extraction. Chromatin was extracted Nuclei from 1×10^9 nuclei of HeLa cells as previously described [8]. Nuclei from 1×10^9 HeLa cells were prepared by pelleting and resuspending cells in buffer containing 3.75 mM Tris at pH 7.5, 20 mM KCl, 0.5 mM EDTA, 0.5 mM DTT, 0.05 mM spermidine, 0.125 mM spermine, 1 mM PMSF and 0.1% digitotin. Cells were homogenized with 10 strokes and nuclei were pelleted at 300*g*. Nuclei were then washed once in wash buffer (20 mM HEPES at pH 7.7, 20 mM KCl, 0.5 mM EDTA, 0.5 mM DTT and 0.5 mM PMSF), followed by wash buffer containing 150 mM NaCl. Nuclei were resuspended in wash

544 buffer supplemented with 150 mM NaCl and 3 mM $CaCl_2$. Chromatin was digested at room
545 temperature using 140 units ml–1 of micrococcal nuclease (Roche, 10107921001) for 35
546 minutes to produce a pool of mono-nucleosomes. Following micrococcal nuclease treatment,
547 extracts were supplemented with 5 mM EGTA and 0.05% NP40 and centrifuged at 10,000g for
548 15 min at 4 °C. The supernatant was then used as the starting material for all
549 immunopurifications.

550 **Affinity purification**. TAP- or LAP -tagged chromatin were purified in two steps. In the first
551 step, native TAP-tagged chromatin was immunoprecipitated by incubating the bulk soluble
552 mono-nucleosome pool with rabbit IgG (Sigma-Aldrich) coupled to Dynabeads M-270 Epoxy
553 (Thermo Fisher Scientific, 14301). Alternatively, CENP-A^{LAP} chromatin was immunoprecipitated
554 using mouse anti-GFP antibody (clones 19C8 and 19F7, Monoclonal Antibody Core Facility at
555 Memorial Sloan-Kettering Cancer Center, New York) [87] coupled to Dynabeads M-270 Epoxy.
556 Chromatin extracts were incubated with antibody bound beads for 16 h at 4 °C. Bound
557 complexes were washed once in buffer A (20 mM HEPES at pH 7.7, 20 mM KCl, 0.4 mM EDTA
558 and 0.4 mM DTT), once in buffer A with 300 mM KCl and finally twice in buffer A with 300 mM
559 KCl, 1 mM DTT and 0.1% Tween 20. In the second step, TAP-chromatin complexes were
560 incubated 16 h in final wash buffer with 50µl recombinant TEV protease, resulting in cleavage
561 of the TAP tag and elution of the chromatin complexes from the beads. Alternatively, CENP-
562 A^{LAP} chromatin was eluted from the beads by cleaving the LAP tag using PreScission protease
563 (4 h, 4°C).

564 **DNA extraction**. Following elution of the chromatin from the beads, Proteinase K (100 µg/ml)
565 was added and samples were incubated for 2 h at 55°C. DNA was purified from proteinase K
566 treated samples using a DNA purification kit following the manufacturer instructions (Promega,

Madison, USA) and was subsequently analyzed either by running a 2% low melting agarose (APEX) gel or by an Agilent 2100 Bioanalyzer by using the DNA 1000 kit. The Bioanalyzer determines the quantity of DNA on the basis of fluorescence intensity.

Quantitative real-time PCR (qPCR). Quantitative real-time PCR (qPCR) was performed using SYBR Green mix (Bio Rad) with CFX384 Bio Rad Real Time System. Primers sequences used in this study: MRGPRE: (forward) 5'-CTGCGCGGATCTCATCTTCC-3' and (reverse) 5'-GGCCCACGATGTAGCAGAA-3'. MMP15: (forward) 5'-GTGCTCGACGAAGAGACCAAG-3' and (reverse) 5'-TTTCACTCGTACCCCGAACTG-3'. HBE1: (forward) 5'-ATGGTGCATTTTACTGCTGAGG-3' and (reverse) 5'-GGGAGACGACAGGTTTCCAAA-3'. Sat2: (forward) 5'-TCGCATAGAATCGAATGGAA-3' and (reverse) 5'-GCATTCGAGTCCGTGGA-3' [16]. α-satellite DNA (from chromosomes 1, 3, 5, 10, 12 and 16): (forward) 5'-CTAGACAGAAGAATTCTCAG-3' and (reverse) 5'-CTGAAATCTCCACTTGC-3' [55]. Melting curve analysis was used to confirm primer specificity. To ensure linearity of the standard curve, reaction efficiencies over the appropriate dynamic range were calculated. Using the dCt method, we calculated fold-enrichment of α-satellite DNA after immunopurification of CENP-A^{TAP} chromatin, compared to its level in the bulk input chromatin. For Repli-seq experiment, we used the dCt method, to calculate fold-enrichment of replicated DNA after immunopurification of BrdU-labeled DNA compared to its level in the bulk input DNA. Reported values are the means of two independent biological replicates with technical duplicates that were averaged for each experiment. Error bars represent SE of the mean.

Immunoblotting. For immunoblot analysis, protein samples were separated by SDS–PAGE, transferred onto PVDF membranes (Millipore) and then probed with the following antibodies: rabbit anti-CENP-A (Cell Signaling, 2186s, 1:1,000), rabbit anti-CENP-B (Millipore, 07-735,

1:200), mouse anti-α-tubulin (Abcam, DM1A, 1:5000), rabbit anti-CAF1p150 (Santa Cruz, sc-10772, 1:500), rabbit anti-CAF1p60 (Bethyl Laboratories, A301-085A, 1:1,000), rabbit anti-CAF1p48 (Bethyl Laboratories, A301-206A, 1:1,000), rabbit anti-MCM2 (Abcam, Ab4461, 1:1,000). Following incubation with HRP-labelled antibody (GE Healthcare, NA931V or NA934V), HRP was detected using enhanced chemiluminescence (ECL) substrate (Thermo Scientific, 34080 or 34096).

Immunofluorescence. $1x10^6$ suspension cells were centrifuged and resuspended with PBS. 10^5 cells were immobilized on glass slide by cytospin centrifugation for 3 min, 800rpm. Cells were then fixed using ice-cold methanol at -20°C for 10 min, followed by washing with cold PBS and then incubated in Triton Block (0.2 M glycine, 2.5% FBS, 0.1% Triton X-100, PBS) for one hour. The following primary antibodies were used: mouse anti-GFP (Roche, 11814460001, 1:500), rabbit anti-CENP-B (Abcam 25734, 1:1,000), human anti-centromere antibodies (ACA, Antibodies Inc, 15-234-0001, 1:500). The following secondary antibodies (Jackson Laboratories) were used for 45 minutes: donkey anti-human TR (1:300), anti-mouse FITC (1:250). TAP fusion proteins were visualized by incubation with FITC-rabbit IgG (Jackson Laboratories, 1:200). Cells were then washed with 0.1% Triton X-100 in PBS, counterstained with DAPI and mounted with mounting medium (Molecular Probes, P36934). Immunofluorescent images were acquired on a Deltavision Core system at x60-100 magnification. 0.2 µm *Z*-stack deconvolved projections were generated using the softWoRx program.

Flow cytometry. Flow cytometry was used to determine the DNA content of the cells as. $1x10^6$ cells were harvested, washed in PBS and fixed in 70% ethanol. Cells were then washed and DNA was stained by incubating cells for 30 min with 1% FBS, 10 µg ml^{-1} propidium iodide and

613 0.25 mg ml^{-1} RNase A in PBS followed by FACS analysis for DNA content using a BD LSR II
614 Flow Cytometer (BD Biosciences).

615 **ChIP-Seq Library Generation and Sequencing.** ChIP libraries were prepared following
616 Illumina protocols with minor modifications (Illumina, San Diego, CA). To reduce biases induced
617 by PCR amplification of a repetitive region, libraries were prepared from 80-100 ng of input or
618 ChIP DNA. The DNA was end-repaired and A-tailed and Illumina Truseq adaptors were ligated.
619 Libraries were run on a 2% agarose gel. Since the chromatin was digested to
620 mononucleosomes, following adaptors ligation the libraries size was 250-280 bp. The libraries
621 were size selected for 200-375 bp. The libraries were then PCR-amplified using only 5-6 PCR
622 cycles since the starting DNA amount was high. Resulting libraries were sequenced using 100
623 bp, paired-end sequencing on a HiSeq 2000 instrument per manufacturer's instructions with
624 some modifications (Illumina, San Diego, CA). Sequence reads are summarized in Table S1.

625 **Initial sequence processing and alignment.** Illumina paired-end reads were merged to
626 determine CENP-A or H3 containing target fragments of varying length using PEAR software
627 [88], with standard parameters (p-value: 0.01, min-overlap: 10 bases, min-assembly length:
628 50bp). Merged paired reads were mapped (Bwa-Mem, standard parameters [89,90]) to the human
629 genome 38 (hg38) assembly (including alternative assemblies), which contain human α-satellite
630 sequence models in each centromeric region ([45]; BioProject: PRJNA193213; [46]). Reads were
631 determined to contain α-satellite if they overlapped sites (BEDTools: intersect [91]) in the genome
632 previously annotated as α-satellite (UCSC table browser [92] was used to obtain a bed file of all
633 sites annotated as ALR/α-Satellite). Additionally, merged sequences were defined as
634 containing α-satellite if they contained an exact match to at least two 18-mers specific to a
635 previously published WGS read database of α-satellite, representing 2.6% of sequences from

636 the HuRef genome [21,48]. Comparisons between the Bwa mapping and 18-mer exact matching
637 based strategies were highly concordant. Total α-satellite DNA content in human genome 38
638 assembly was estimated by using the UCSC RepeatMasker Annotation [92,93]. Summary of reads
639 obtained is shown in Table S1.

640 **ChIP-seq peak calling.** Enrichment peaks for ChIP experiments were determined using SICER
641 algorithm (v1.0.3) [50] using relevant input reads as background, with stringent parameters
642 previously optimized for human CENP-A [36]: threshold for redundancy allowed for chip reads: 1,
643 threshold for redundancy allowed for control reads: 1, window size: 200 bps, fragment size: 150
644 bps, shift in window length is 150, effective genome size as a fraction of the reference genome
645 of hg38: 0.74, gap size: 400 bps, e-value for identification of candidate islands that exhibit
646 clustering: 1000, and false discovery rate controlling significance: 0.00001. In parallel, MACs
647 peak calling was performed (macs14) [49], and wiggle tracks were created to represent read
648 depth of each dataset independently. Finally, we performed a final, rigorous evaluation of
649 ectopic CENP-A peaks, or peaks predicted outside of centromeric regions, using k-mer
650 enrichment (previously described [21]). Each ectopic peak was reformatted into 50-bp sliding
651 windows (in both orientations, with slide of 1bp). The normalized frequency of each 50-mer
652 candidate windows were evaluated in each ChIP-seq dataset relative to a normalized observed
653 frequency in the corresponding background dataset. Scores were determined as the log
654 transformed normalized value of the ratio between ChIP-seq and background, and those with
655 a score greater than or equal to 2 were included in our study as a high-confident enrichment
656 set.

657 **Analysis of CENP-A peaks overlap with functional annotation.** Ectopic CENP-A peak calls,
658 i.e. those that did not overlap with centromeric α-satellite DNA, were evaluated for enrichment

with functional annotation if they were supported between replicate ChIP-seq experiments and overlapped at least one enriched 50mer with a log-transformed normalized ratio >=2, or with a minimum standard ratio of 5-fold. Resulting high-confident ectopic peak calls were intersected (BEDTools: intersect; [91] with select functional datasets in the genome (UCSC table browser, [92]. Peaks that intersect with GRCh38 RefSeq genes (including introns and exons) w/ minimum overlap (-f 0.9;or 90%) required as a fraction of SICER peaks, as well as 1000 bp upstream and downstream (with minimum overlap of 1bp with SICER peak). To evaluate the role of expression, gene annotation was catalogued further based on Intersection with ENCODE HeLa expression data (wgEncodeRegTxnCshlLongRnaSeqHelas3CellPapRawSigPooled) with RefSeq gene annotations (where 22,211 RefSeq Genes (40.5% of total) demonstrated at least >=10 average reads/gene; and highly expressed RefSeq Genes (10,033, or 18.3% of total) RefSeq genes are defined as >=100 average reads/gene). To investigate peak overlap with sites of CTCF enrichment, we intersected peaks with two ENCODE replicate datasets: HeLa-S3.CTCF_Ht1.bed and HeLa-S3.CTCF_Ht2.bed (with minimum overlap of 20 bp). To study the overlap with sites of open chromatin peaks were intersected with ENCODE datasets: HeLa-S3.UW_DNase1_HS.Ht1.bed and HeLa-S3.UW_DNase1_HS.Ht2.bed with minimum overlap of 100 bases. Results were evaluated relative to a simulated peak dataset to test if observed peak counts are higher than expected by chance. Simulations were repeated 100x to provide basic summary statisics: average, standard deviation, max/min, relative enrichment value and empirical p-value.

Repli-seq experiments. BrdU labeled DNA across S phase was prepared as previously described [56] with some modifications. Briefly, cells were synchronized using double thymidine block and release [8]. Following release from double thymidine block, cells were labeled with BrdU (Sigma, B5002) for 1 hour by adding BrdU to the culture medium to a final concentration

683 of 50 µM. For labeling at early-S (S1), BrdU was added immediately after release (S0). For
684 labeling at mid-S (S4), BrdU was added 3 hours after release (S3). For labeling at late-S (S7),
685 BrdU was added 6 hours after release (S6). Following labeling with BrdU, genomic DNA was
686 extracted, sonicated and heat denaturated as previously described [56]. BrdU labeled DNA was
687 immunoprecipitated using an anti-BrdU antibody (Becton-Dickinson Biosciences, 555627)
688 coupled to magnetic Dyna M-270 epoxy beads (Thermo Fisher Scientific, 14301). Eluted single-
689 stranded DNA was made into double stranded DNA using random-prime extension (Thermo
690 Fisher Scientific, Random Primers DNA Labeling Kit, 18187-013). Following cleanup of the
691 double stranded DNA (QIAgen QiaQuick PCR Purification Kit, 28104), the DNA was validated
692 by performing quantitative real-time PCR using primers for MRGPRE, MMP15, HBE1 and Sat2.
693 Libraries were then prepared as described above, and sequenced using the Illumina instrument
694 per manufacturer's instructions (Illumina, San Diego, CA) with the exception that following
695 adapter ligation, repli-seq libraries were size selected between 250-500bp.

696 **Mass spectrometry identification of proteins associating with CENP-A^{TAP} chromatin.**
697 CENP-A^{TAP} was immunoprecipitated from the chromatin fraction of randomly cycling cells or
698 late S synchronized cells as described above. Following beads washes, beads were snap
699 frozen in liquid nitrogen. Samples were diluted using 100 mM Tris pH 8.5 to a final concentration
700 of 2 M urea and digested with trypsin (Promega) overnight at 37 degrees Celsius. The protein
701 digests were pressure-loaded onto 250 micron i.d. fused silica capillary (Polymicro
702 Technologies) columns with a Kasil frit packed with 3 cm of 5 micron Partisphere strong cation
703 exchange (SCX) resin (Whatman) and 3 cm of 5 micron C18 resin (Phenomenex). After
704 desalting, each bi-phasic column was connected to a 100 micron i.d. fused silica capillary
705 (Polymicro Technologies) analytical column with a 5 micron pulled-tip, packed with 10 cm of 5
706 micron C18 resin (Phenomenex). Each MudPIT column were placed inline with an 1200

quaternary HPLC pump (Agilent Technologies) and the eluted peptides were electrosprayed directly into an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific). The buffer solutions used were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B) and 500 mM ammonium acetate/5% acetonitrile/0.1% formic acid (buffer C). A ten-step MudPIT, each step consisting of a 120 minute elution gradient, was run with salt pulses of 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70% and 100% buffer C and 90% buffer C/10% buffer B. The MS/MS cycle consisted of one full scan mass spectrum (400-1600 m/z) at 60 K resolution followed by five data-dependent collision induced dissocation (CID) MS/MS spectra. Charge state exclusion was enabled with +1 and unassigned charge states rejected for fragmentation. Application of mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system (Thermo Scientific). MS/MS spectra were extracted using RawXtract (version 1.9.9) [94]. MS/MS spectra were searched with the ProLuCID algorithm [95] against a human UniProt protein database downloaded on 03-25-2014 that had been supplemented with common contaminants and concatenated to a decoy database in which the sequence for each entry in the original database was reversed [96]. The ProLuCID search was performed using full enzyme specificity (cleavage C-terminal to Arg or Lys residue). The data was searched using a precursor mass tolerance of 50 ppm and a fragment ion mass tolerance of 600 ppm. The ProLuCID search results were assembled and filtered using the DTASelect (version 2.0) algorithm [97]. DTASelect assesses the validity of peptide-spectra matches using the cross-correlation score (XCorr) and normalized difference in cross-correlation scores (deltaCN). The search results are grouped by charge state and tryptic status and each sub-group is analyzed by discriminant analysis based on a non-parametric fit of the distribution of forward and reversed matches. A minimum of two peptides was required for each protein identification. All peptide-spectra matches had less than 10 ppm mass error. The protein false

positive rate was below one percent for all experiments.

Quantification and statistical analysis. For all experiments shown, n is indicated in the figure legends. Values represent the mean ± s.e.m (as indicated in the figure legends).

Data availability. The datasets generated during the current study were deposited at GEO under primary accession number GSE111381.

The datasets generated during the current study were compared to the following publicly available datasets: ENCODE HeLa DNase-seq (DNase-1 hyper sensitive sites; GEO accsession: GSE90432), ENCODE HeLa CTCF ChIP-sequencing (GEO accession: GSM749729 and GSM749739), ENCODE HeLa H3K27me3 ChIP-sequencing (GEO accession: GSM945208), ENCODE HeLa expression data (UCSC Accession: wgEncodeEH000130), ENCODE HeLa-S3 Repli-seq (GEO accession: GSM923449).

Author Contributions

Y.N-A. and D.W.C. conceived and designed experiments and wrote the manuscript. Y.N-A performed experiments. KH.M analyzed the sequencing data. M.A.M. and O.S. analyzed data and performed experiments. D.F. suggested experiments and provided key experimental input. A.Y.L and B.R prepared sequencing libraries and provided resources. A.A and J.YIII performed mass spectrometry experiments and provided resources.

Acknowledgments

The authors would like to thank A. Desai, P. Ly and C. Eissler for critical discussion and helpful suggestions, L.E.T Jansen (Gulbenkian Institute, Oeiras, Portugal) for providing reagents. This work was supported by grants (R01 GM-074150 and R35 GM-122476) from the National Institutes of Health to D.W.C., who receives salary support from the Ludwig Institute for Cancer Research.

Conflict of Interests

The authors declare that they have no conflict of interest.

References

1. Wevrick, R. & Willard, H.F. Long-range organization of tandem arrays of alpha satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc Natl Acad Sci U S A* **86**, 9394-8 (1989).

2. Cleveland, D.W., Mao, Y. & Sullivan, K.F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407-21 (2003).

3. Willard, H.F. Chromosome-specific organization of human alpha satellite DNA. *American journal of human genetics* **37**, 524-32 (1985).

4. Manuelidis, L. & Wu, J.C. Homology between human and simian repeated DNA. *Nature* **276**, 92-4 (1978).

5. Earnshaw, W.C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* **91**, 313-21 (1985).

6. Palmer, D.K., O'Day, K., Wener, M.H., Andrews, B.S. & Margolis, R.L. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J Cell Biol* **104**, 805-15 (1987).

7. Bodor, D.L. *et al.* The quantitative architecture of centromeric chromatin. *eLife* **3**, e02137 (2014).

8. Nechemia-Arbely, Y. *et al.* Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. *J Cell Biol* **216**, 607-621 (2017).

9. Sullivan, B.A. & Karpen, G.H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature structural & molecular biology* **11**, 1076-83 (2004).

10. Sullivan, L.L., Boivin, C.D., Mravinac, B., Song, I.Y. & Sullivan, B.A. Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. *Chromosome Res* **19**, 457-70 (2011).

11. Karpen, G.H. & Allshire, R.C. The case for epigenetic effects on centromere identity and function. *Trends in genetics : TIG* **13**, 489-96 (1997).

12. Stimpson, K.M. & Sullivan, B.A. Epigenomics of centromere assembly and function. *Current opinion in cell biology* **22**, 772-80 (2010).

13. Marshall, O.J., Chueh, A.C., Wong, L.H. & Choo, K.H. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* **82**, 261-82 (2008).

14. Okada, T. *et al.* CENP-B controls centromere formation depending on the chromatin context. *Cell* **131**, 1287-300 (2007).

15. Ohzeki, J., Nakano, M., Okada, T. & Masumoto, H. CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol* **159**, 765-75 (2002).

16. Ohzeki, J. *et al.* Breaking the HAC Barrier: histone H3K9 acetyl/methyl balance regulates CENP-A assembly. *EMBO J* **31**, 2391-402 (2012).

17. Kouprina, N. *et al.* Cloning of human centromeres by transformation-associated recombination in yeast and generation of functional human artificial chromosomes. *Nucleic Acids Res* **31**, 922-34 (2003).

18. Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K. & Willard, H.F. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* **15**, 345-55 (1997).

19. Grimes, B.R., Rhoades, A.A. & Willard, H.F. Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. *Mol Ther* **5**, 798-805 (2002).

20. Maloney, K.A. *et al.* Functional epialleles at an endogenous human centromere. *Proc Natl Acad Sci U S A* **109**, 13704-9 (2012).

21. Hayden, K.E. *et al.* Sequences associated with centromere competency in the human genome. *Molecular and cellular biology* **33**, 763-72 (2013).

22. Fachinetti, D. *et al.* A two-step mechanism for epigenetic specification of centromere identity and function. *Nature cell biology* **15**, 1056-66 (2013).

23. Foltz, D.R. *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nature cell biology* **8**, 458-69 (2006).

24. Okada, M. *et al.* The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat Cell Biol* **8**, 446-57 (2006).

25. Hori, T. *et al.* CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* **135**, 1039-52 (2008).

26. Hori, T., Shang, W.H., Takeuchi, K. & Fukagawa, T. The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. *The Journal of cell biology* **200**, 45-60 (2013).

27. Padeganeh, A. *et al.* Octameric CENP-A nucleosomes are present at human centromeres throughout the cell cycle. *Current biology : CB* **23**, 764-9 (2013).

28. Hasson, D. *et al.* The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature structural & molecular biology* **20**, 687-95 (2013).

29. Jansen, L.E., Black, B.E., Foltz, D.R. & Cleveland, D.W. Propagation of centromeric chromatin requires exit from mitosis. *The Journal of cell biology* **176**, 795-805 (2007).

30. Mellone, B.G. *et al.* Assembly of Drosophila centromeric chromatin proteins during mitosis. *PLoS genetics* **7**, e1002068 (2011).

31. Schuh, M., Lehner, C.F. & Heidmann, S. Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. *Curr Biol* **17**, 237-43 (2007).

32. Nechemia-Arbely, Y., Fachinetti, D. & Cleveland, D.W. Replicating centromeric chromatin: spatial and temporal control of CENP-A assembly. *Experimental cell research* **318**, 1353-60 (2012).

33. Foltz, D.R. *et al.* Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell* **137**, 472-84 (2009).

34. Dunleavy, E.M. *et al.* HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485-97 (2009).

35. Silva, M.C. *et al.* Cdk activity couples epigenetic centromere inheritance to cell cycle progression. *Developmental Cell* **22**, 52-63 (2012).

36. Lacoste, N. *et al.* Mislocalization of the centromeric histone variant CenH3/CENP-A in human cells depends on the chaperone DAXX. *Molecular cell* **53**, 631-44 (2014).

37. Van Hooser, A.A. *et al.* Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *Journal of cell science* **114**, 3529-42 (2001).

38. Shrestha, R.L. *et al.* Mislocalization of centromeric histone H3 variant CENP-A contributes to chromosomal instability (CIN) in human cells. *Oncotarget* **8**, 46781-46800 (2017).

39. Filipescu, D. *et al.* Essential role for centromeric factors following p53 loss and oncogenic transformation. *Genes Dev* **31**, 463-480 (2017).

40. Heun, P. *et al.* Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Developmental Cell* **10**, 303-15 (2006).

41. Au, W.C., Crisp, M.J., DeLuca, S.Z., Rando, O.J. & Basrai, M.A. Altered dosage and mislocalization of histone H3 and Cse4p lead to chromosome loss in Saccharomyces cerevisiae. *Genetics* **179**, 263-75 (2008).

42. Collins, K.A., Camahort, R., Seidel, C., Gerton, J.L. & Biggins, S. The overexpression of a Saccharomyces cerevisiae centromeric histone H3 variant mutant protein leads to a defect in kinetochore biorientation. *Genetics* **175**, 513-25 (2007).

43. Mata, J.F., Lopes, T., Gardner, R. & Jansen, L.E. A rapid FACS-based strategy to isolate human gene knockin and knockout clones. *PloS one* **7**, e32646 (2012).

44. Conde e Silva, N. *et al.* CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. *Journal of molecular biology* **370**, 555-73 (2007).

45. Miga, K.H. *et al.* Centromere reference models for human chromosomes X and Y satellite arrays. *Genome research* **24**, 697-707 (2014).

865 46. Miga, K.H. *et al.* Centromeric reference models for the 22 human autosomes.
866 (unpublished data).

867 47. Schneider, V.A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies
868 demonstrates the enduring quality of the reference assembly. *Genome Res* **27**, 849-864
869 (2017).

870 48. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS biology* **5**,
871 e254 (2007).

872 49. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137
873 (2008).

874 50. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone
875 modification ChIP-Seq data. *Bioinformatics* **25**, 1952-8 (2009).

876 51. Earnshaw, W.C. *et al.* Molecular cloning of cDNA for CENP-B, the major human
877 centromere autoantigen. *J Cell Biol* **104**, 817-29 (1987).

878 52. Amor, D.J. & Choo, K.H. Neocentromeres: role in human disease, evolution, and
879 centromere study. *American journal of human genetics* **71**, 695-714 (2002).

880 53. Amor, D.J. *et al.* Human centromere repositioning "in progress". *Proceedings of the*
881 *National Academy of Sciences of the United States of America* **101**, 6542-7 (2004).

882 54. Hasson, D. *et al.* Formation of novel CENP-A domains on tandem repetitive DNA and
883 across chromosome breakpoints on human chromosome 8q21 neocentromeres.
884 *Chromosoma* **120**, 621-32 (2011).

885 55. Alonso, A. *et al.* Co-localization of CENP-C and CENP-H to discontinuous domains of
886 CENP-A chromatin at human neocentromeres. *Genome biology* **8**, R148 (2007).

887 56. Hansen, R.S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in
888 human replication timing. *Proc Natl Acad Sci U S A* **107**, 139-44 (2010).

889 57. Ryba, T., Battaglia, D., Pope, B.D., Hiratani, I. & Gilbert, D.M. Genome-scale analysis of
890 replication timing: from bench to bioinformatics. *Nat Protoc* **6**, 870-95 (2011).

891 58. Hassan, K.M., Norwood, T., Gimelli, G., Gartler, S.M. & Hansen, R.S. Satellite 2
892 methylation patterns in normal and ICF syndrome cells and association of
893 hypomethylation with advanced replication. *Hum Genet* **109**, 452-62 (2001).

894 59. Bui, M. *et al.* Cell-cycle-dependent structural transitions in the human CENP-A
895 nucleosome in vivo. *Cell* **150**, 317-26 (2012).

896 60. O'Keefe, R.T., Henderson, S.C. & Spector, D.L. Dynamic organization of DNA replication
897 in mammalian cell nuclei: spatially and temporally defined replication of chromosome-
898 specific alpha-satellite DNA sequences. *J Cell Biol* **116**, 1095-110 (1992).

61. Shelby, R.D., Monier, K. & Sullivan, K.F. Chromatin assembly at kinetochores is uncoupled from DNA replication. *J Cell Biol* **151**, 1113-8 (2000).

62. Ten Hagen, K.G., Gilbert, D.M., Willard, H.F. & Cohen, S.N. Replication timing of DNA sequences associated with human centromeres and telomeres. *Mol Cell Biol* **10**, 6348-55 (1990).

63. Erliandri, I. *et al.* Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. *Nucleic Acids Res* **42**, 11502-16 (2014).

64. Nagpal, H. *et al.* Dynamic changes in CCAN organization through CENP-C during cell-cycle progression. *Mol Biol Cell* **26**, 3768-76 (2015).

65. Screpanti, E. *et al.* Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr Biol* **21**, 391-8 (2011).

66. Nishino, T. *et al.* CENP-T provides a structural platform for outer kinetochore assembly. *EMBO J* **32**, 424-36 (2013).

67. Guse, A., Carroll, C.W., Moree, B., Fuller, C.J. & Straight, A.F. In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* **477**, 354-8 (2011).

68. Carroll, C.W., Milks, K.J. & Straight, A.F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J Cell Biol* **189**, 1143-55 (2010).

69. Petrovic, A. *et al.* Structure of the MIS12 Complex and Molecular Basis of Its Interaction with CENP-C at Human Kinetochores. *Cell* **167**, 1028-1040 e15 (2016).

70. Rago, F., Gascoigne, K.E. & Cheeseman, I.M. Distinct organization and regulation of the outer kinetochore KMN network downstream of CENP-C and CENP-T. *Curr Biol* **25**, 671-7 (2015).

71. Shono, N. *et al.* CENP-C and CENP-I are key connecting factors for kinetochore and CENP-A assembly. *J Cell Sci* **128**, 4572-87 (2015).

72. Klare, K. *et al.* CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J Cell Biol* **210**, 11-22 (2015).

73. Kato, H. *et al.* A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**, 1110-3 (2013).

74. McKinley, K.L. *et al.* The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. *Mol Cell* **60**, 886-98 (2015).

75. Weir, J.R. *et al.* Insights from biochemical reconstitution into the architecture of human kinetochores. *Nature* **537**, 249-253 (2016).

76. Hoffmann, S. *et al.* CENP-A Is Dispensable for Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly. *Cell Rep* **17**, 2394-2404 (2016).

77. Smith, S. & Stillman, B. Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J* **10**, 971-80 (1991).

78. Krude, T. Chromatin assembly during DNA replication in somatic cells. *Eur J Biochem* **263**, 1-5 (1999).

79. Verreault, A., Kaufman, P.D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95-104 (1996).

80. Shang, W.H. *et al.* Acetylation of histone H4 lysine 5 and 12 is required for CENP-A deposition into centromeres. *Nat Commun* **7**, 13465 (2016).

81. Hayashi, T. *et al.* Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* **118**, 715-29 (2004).

82. Kaufman, P.D., Kobayashi, R., Kessler, N. & Stillman, B. The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell* **81**, 1105-14 (1995).

83. Huang, H. *et al.* A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks. *Nat Struct Mol Biol* **22**, 618-26 (2015).

84. Gascoigne, K.E. *et al.* Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* **145**, 410-22 (2011).

85. Zhang, W. *et al.* Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nature communications* **7**, 12619 (2016).

86. Sun, X. *et al.* Elevated expression of the centromere protein-A(CENP-A)-encoding gene as a prognostic and predictive biomarker in human cancers. *Int J Cancer* **139**, 899-907 (2016).

87. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738-48 (2008).

88. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-20 (2014).

89. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010).

90. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *eprint arXiv* (2013).

91. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).

968 92. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic acids research* **32**, D493-6 (2004).

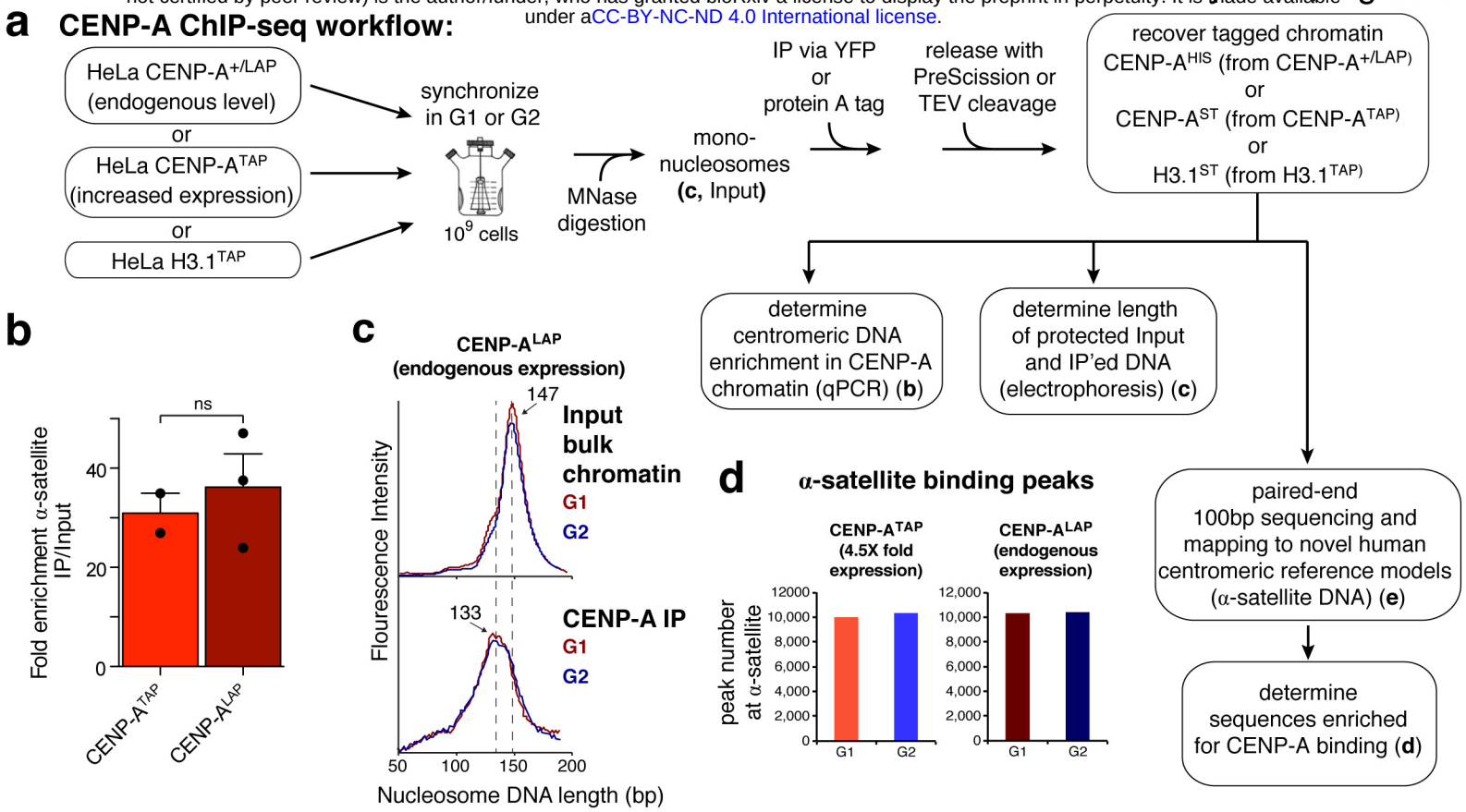
970 93. Rosenbloom, K.R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic acids research* **43**, D670-81 (2015).

972 94. McDonald, W.H. *et al.* MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom* **18**, 2162-8 (2004).

975 95. Xu, T. *et al.* ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J Proteomics* **129**, 16-24 (2015).

977 96. Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. & Gygi, S.P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**, 43-50 (2003).

981 97. Tabb, D.L., McDonald, W.H. & Yates, J.R., 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1**, 21-6 (2002).
984



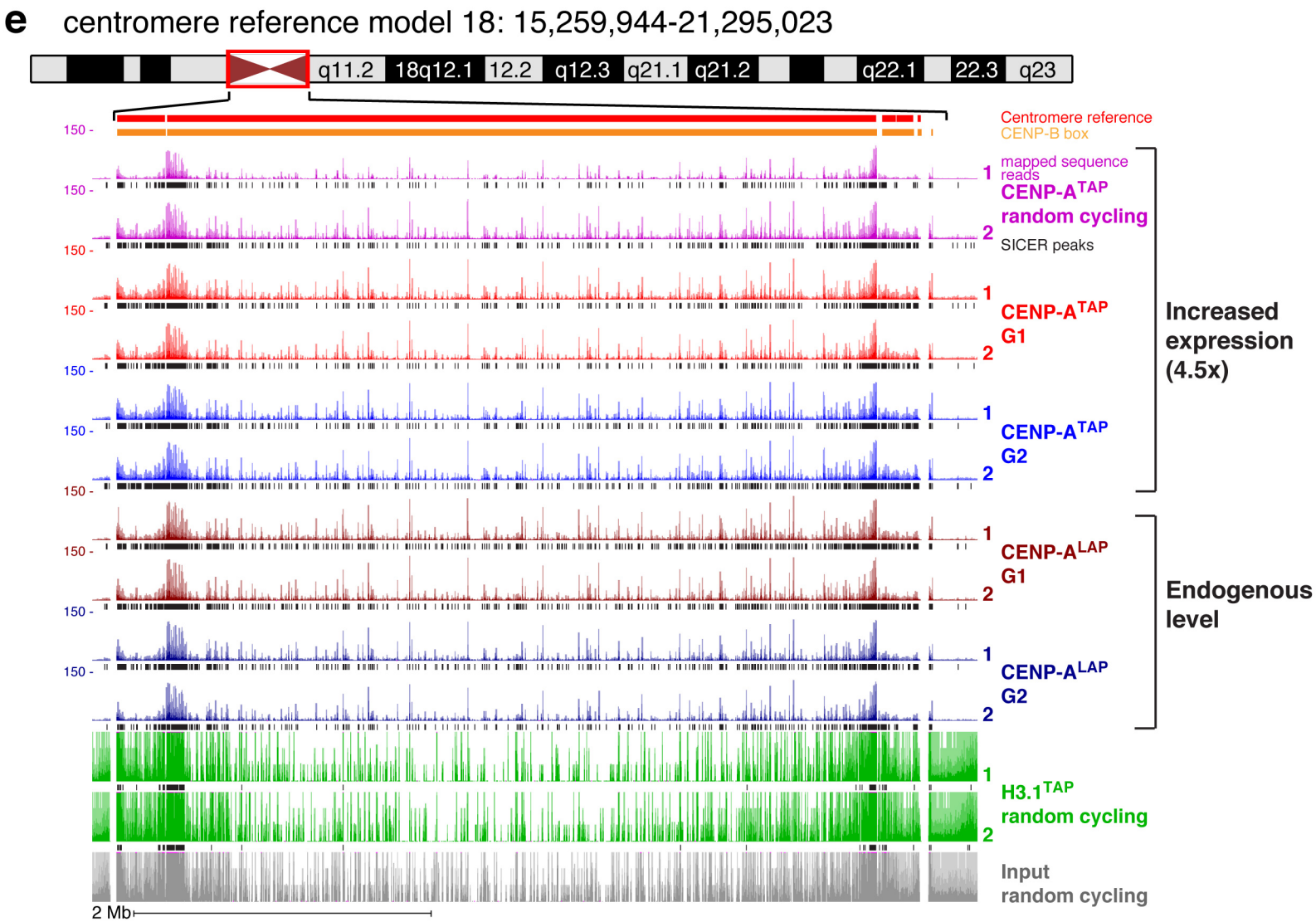


Fig 1. CENP-A ChIP-seq identifies CENP-A binding at reference centromeres of 23 human chromosomes

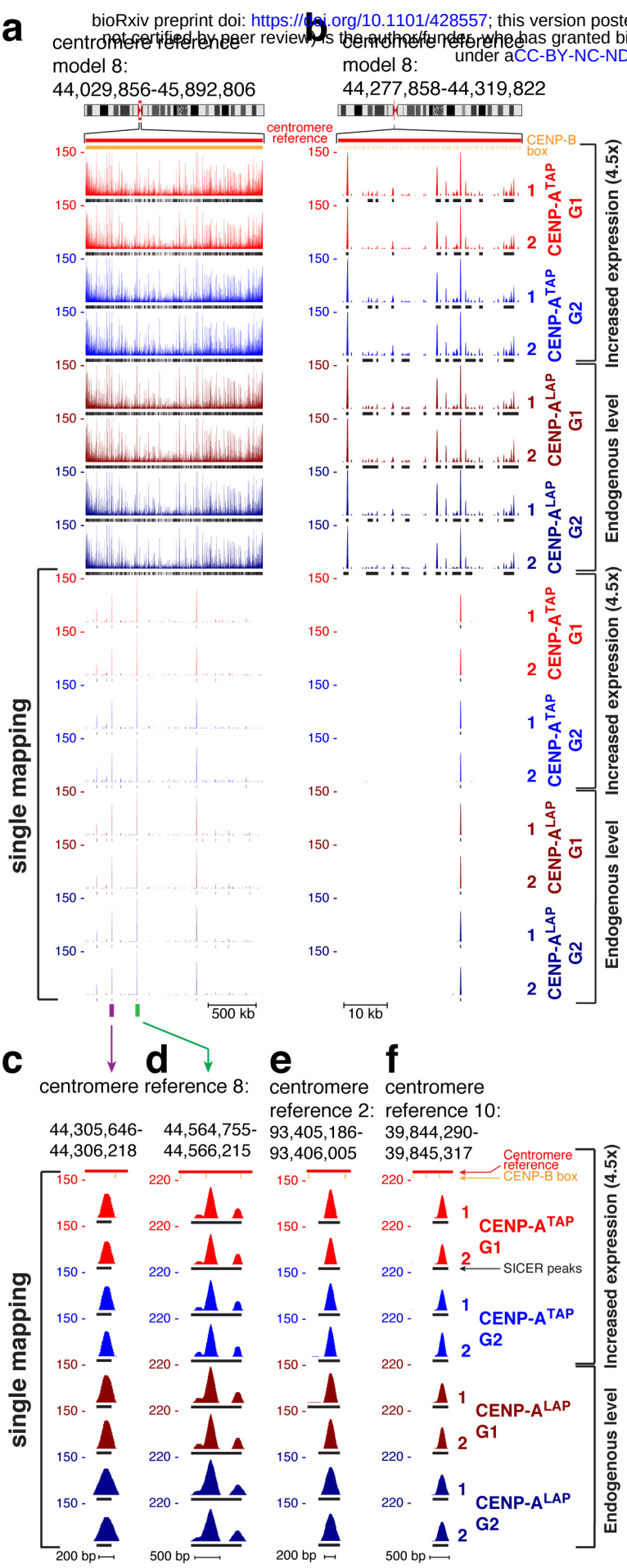


Figure 2. Retention of centromeric CENP-A through DNA replication

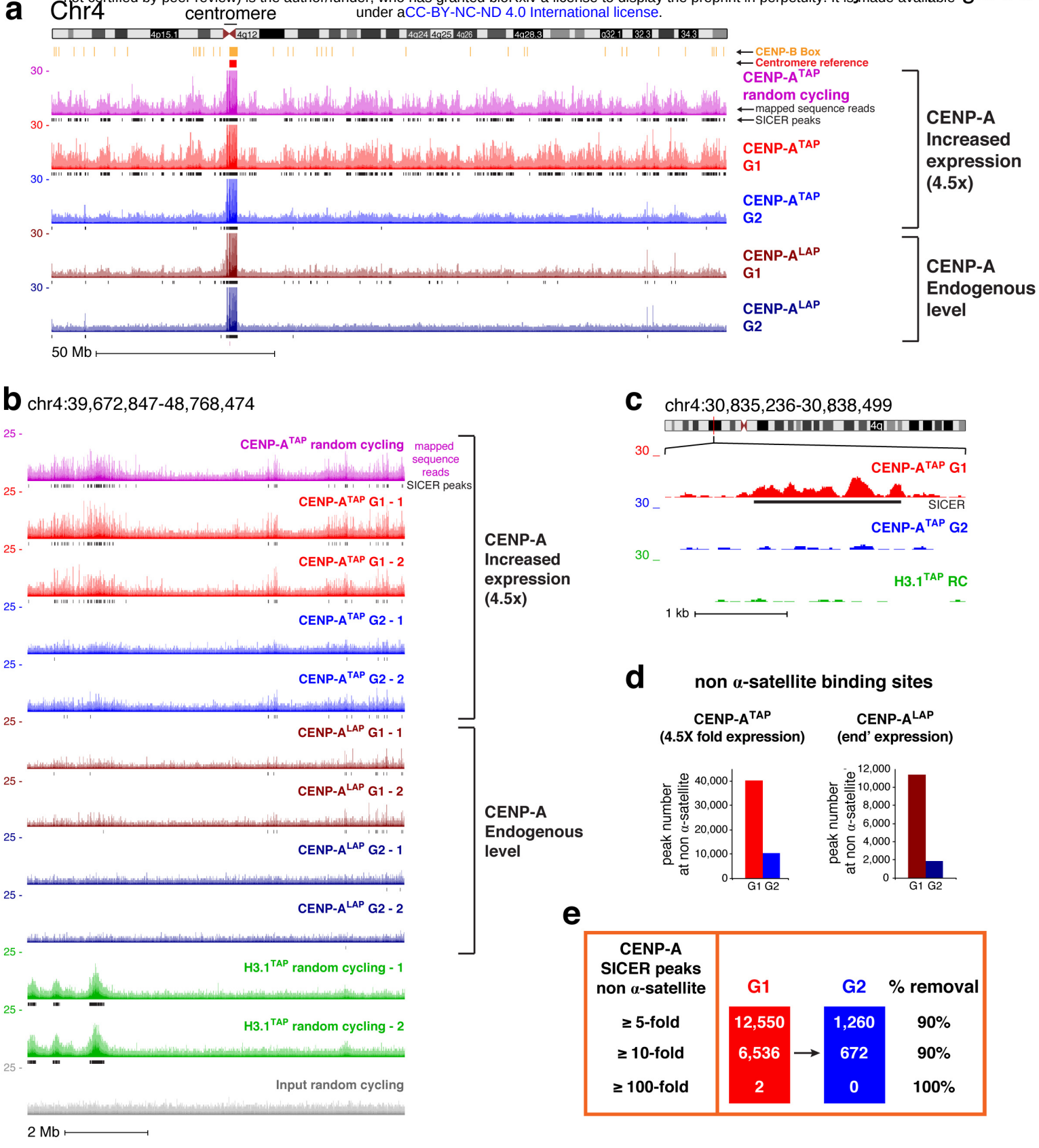


Figure 3. Sites of CENP-A assembly onto chromosome arms in early G1 are removed by G2

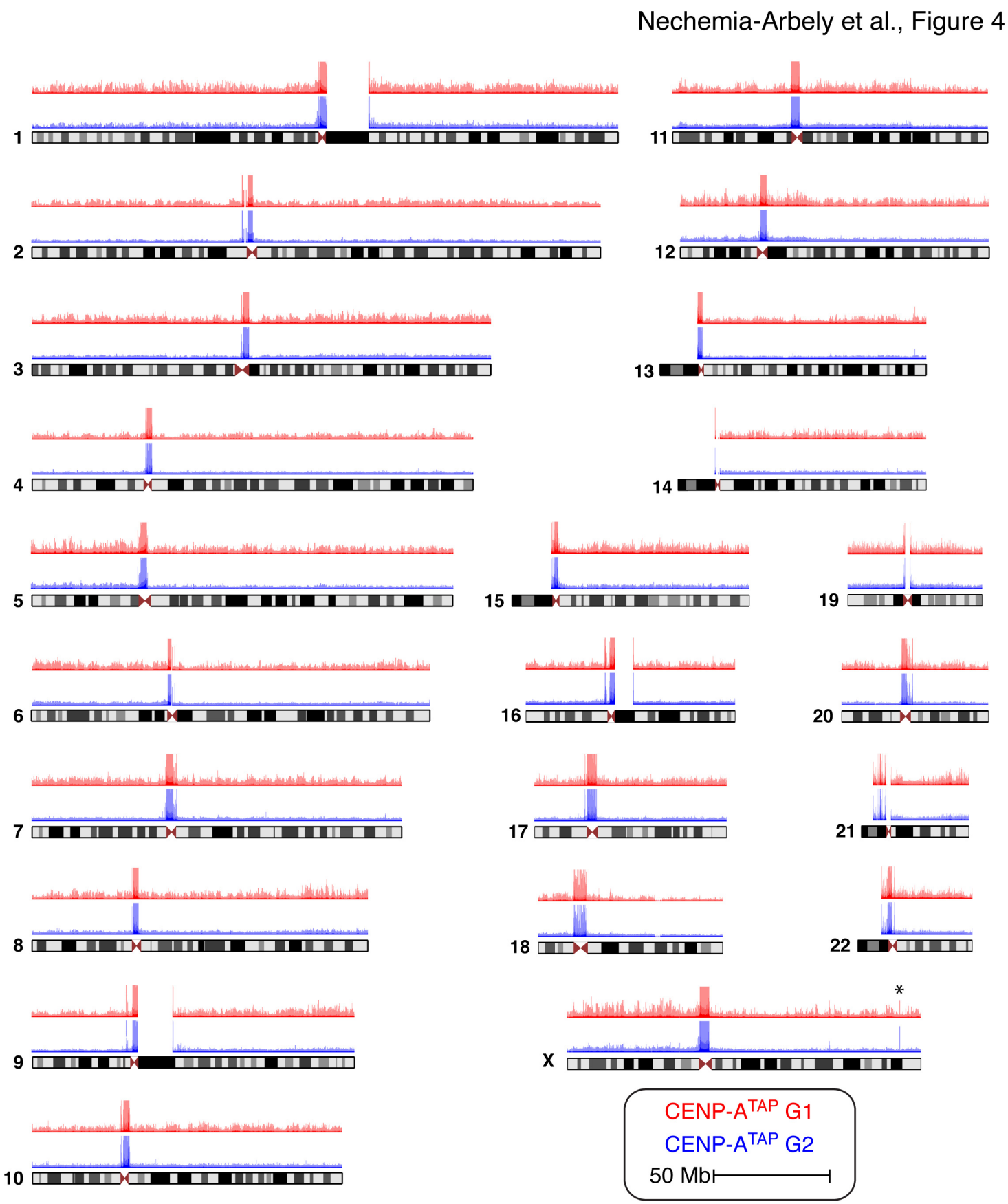


Fig. 4. Ectopic CENP-A is removed following DNA replication from the arms of all 23 human chromosomes.

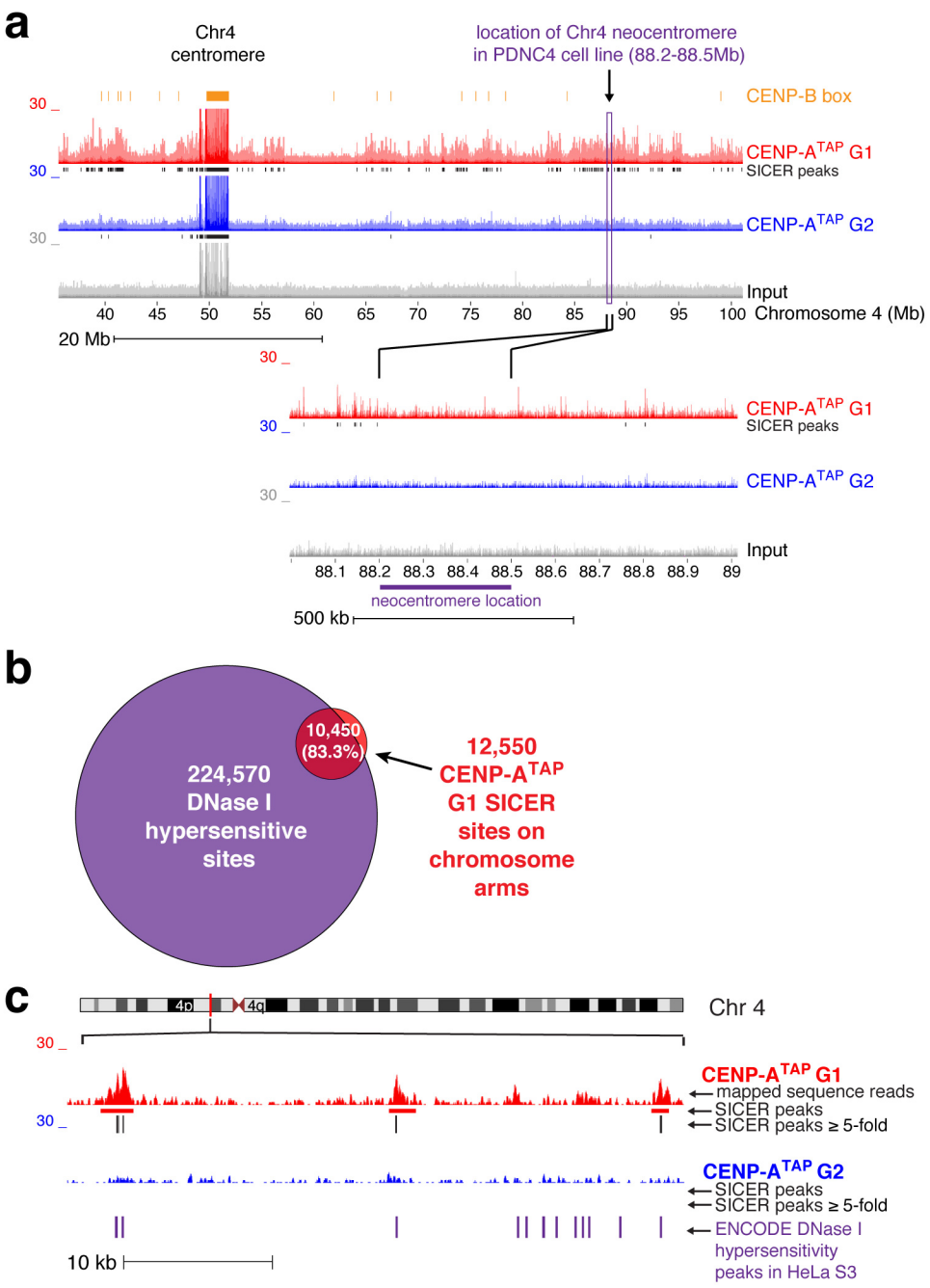
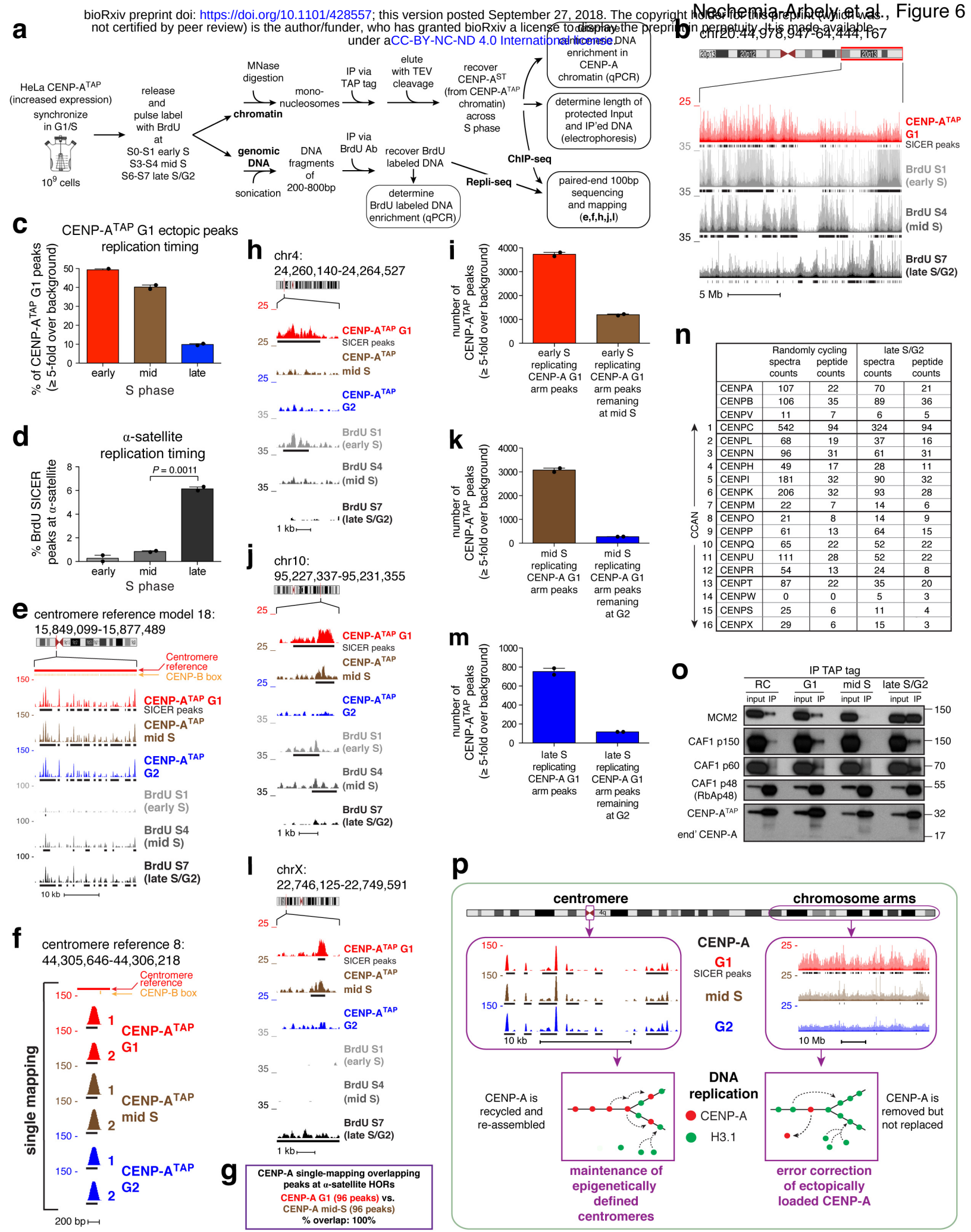


Fig 5. Sites of deposition of CENP-A on chromosome arms are not seeding hotspots for neocentromere formation.



		Randomly cycling spectra counts	Randomly cycling peptide counts	late S/G2 spectra counts	late S/G2 peptide counts
	CENPA	107	22	70	21
	CENPB	106	35	89	36
	CENPV	11	7	6	5
1	CENPC	542	94	324	94
2	CENPL	68	19	37	16
3	CENPN	96	31	61	31
4	CENPH	49	17	28	11
5	CENPI	181	32	90	32
6	CENPK	206	32	93	28
7	CENPM	22	7	14	6
8	CENPO	21	8	14	9
9	CENPP	61	13	64	15
10	CENPQ	65	22	52	22
11	CENPU	111	28	52	22
12	CENPR	54	13	24	8
13	CENPT	87	22	35	20
14	CENPW	0	0	5	3
15	CENPS	25	6	11	4
16	CENPX	29	6	15	3

Figure 6. Ectopic CENP-A is removed contemporaneously with replication fork progression, while centromeric CENP-A is retained

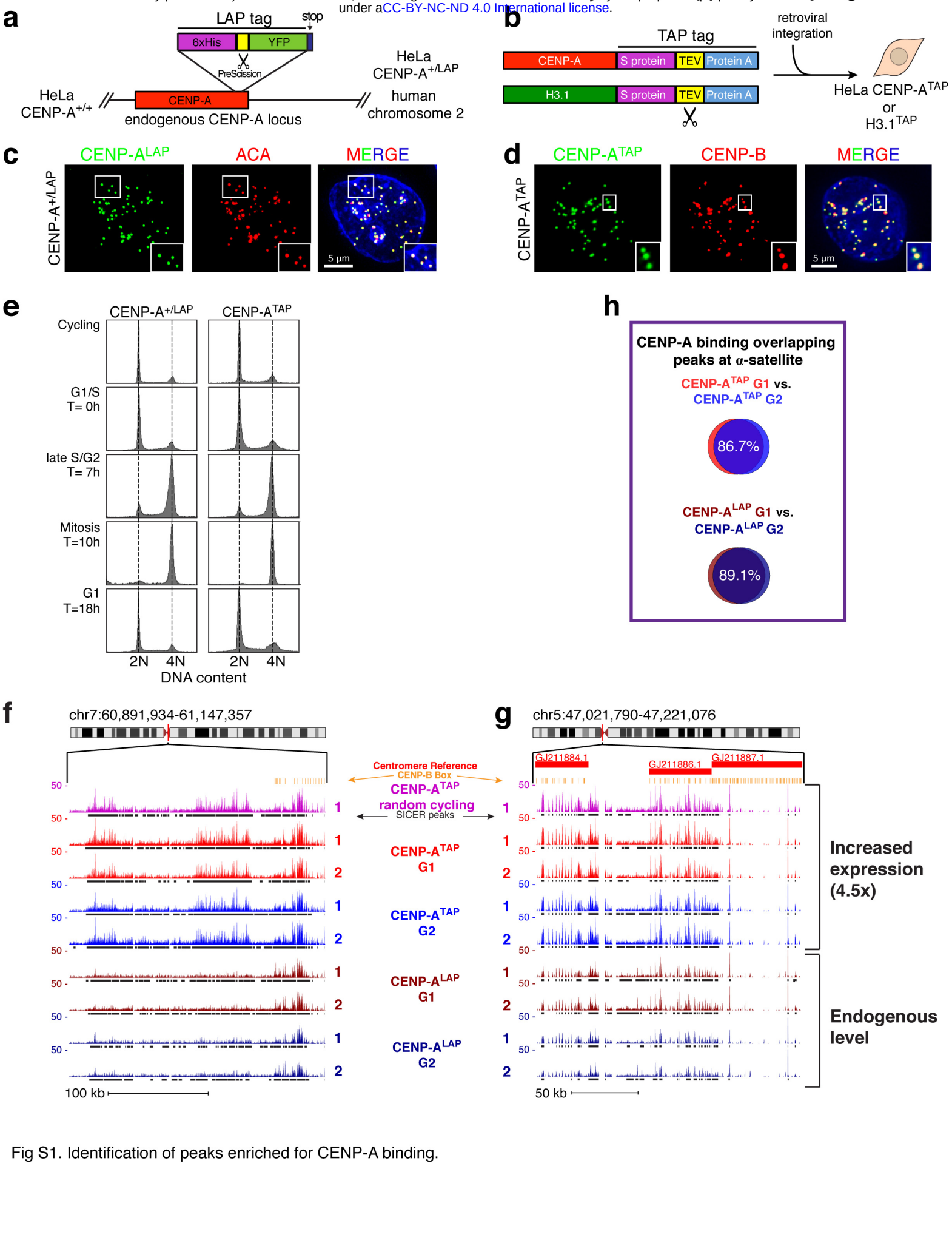
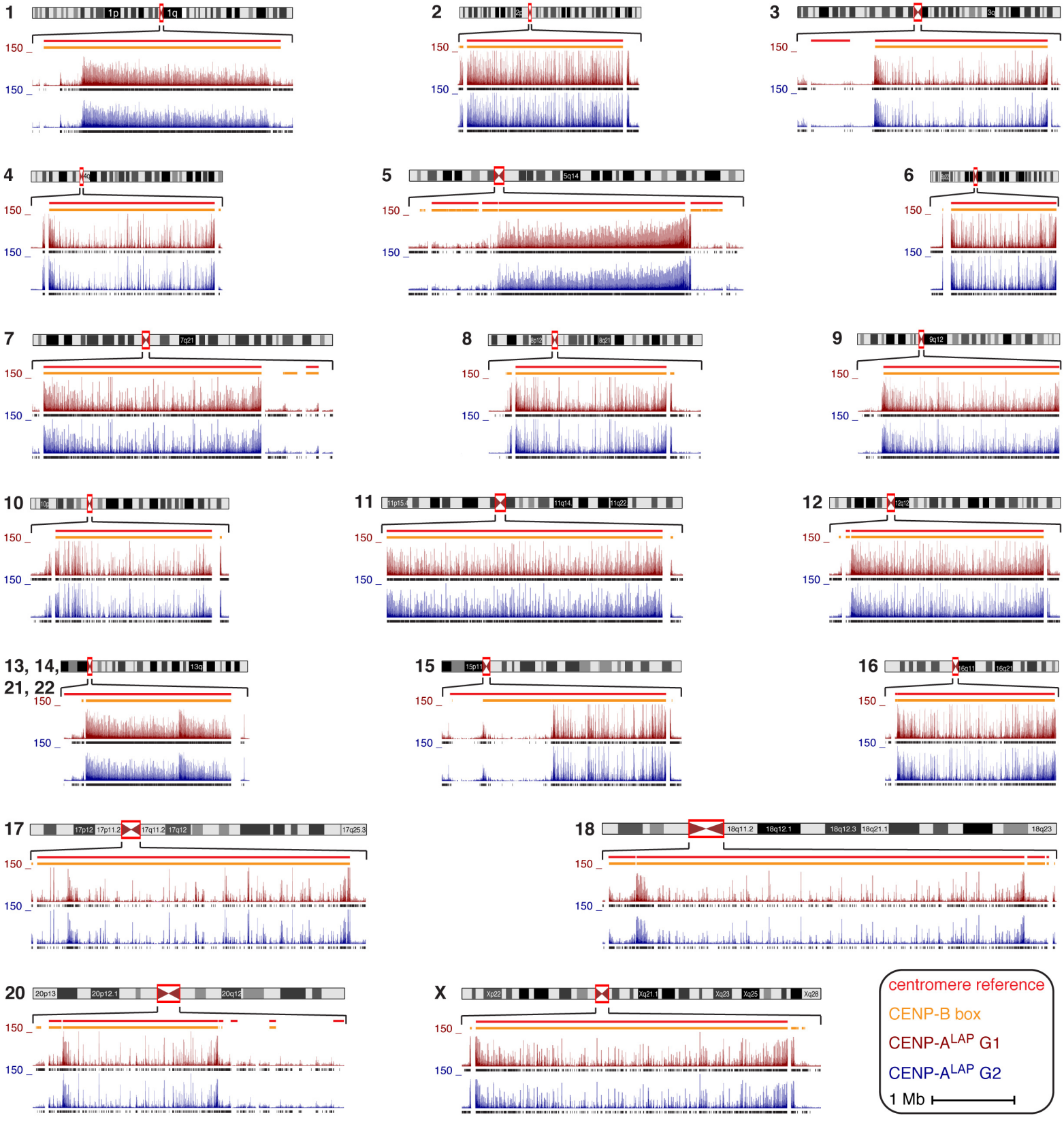
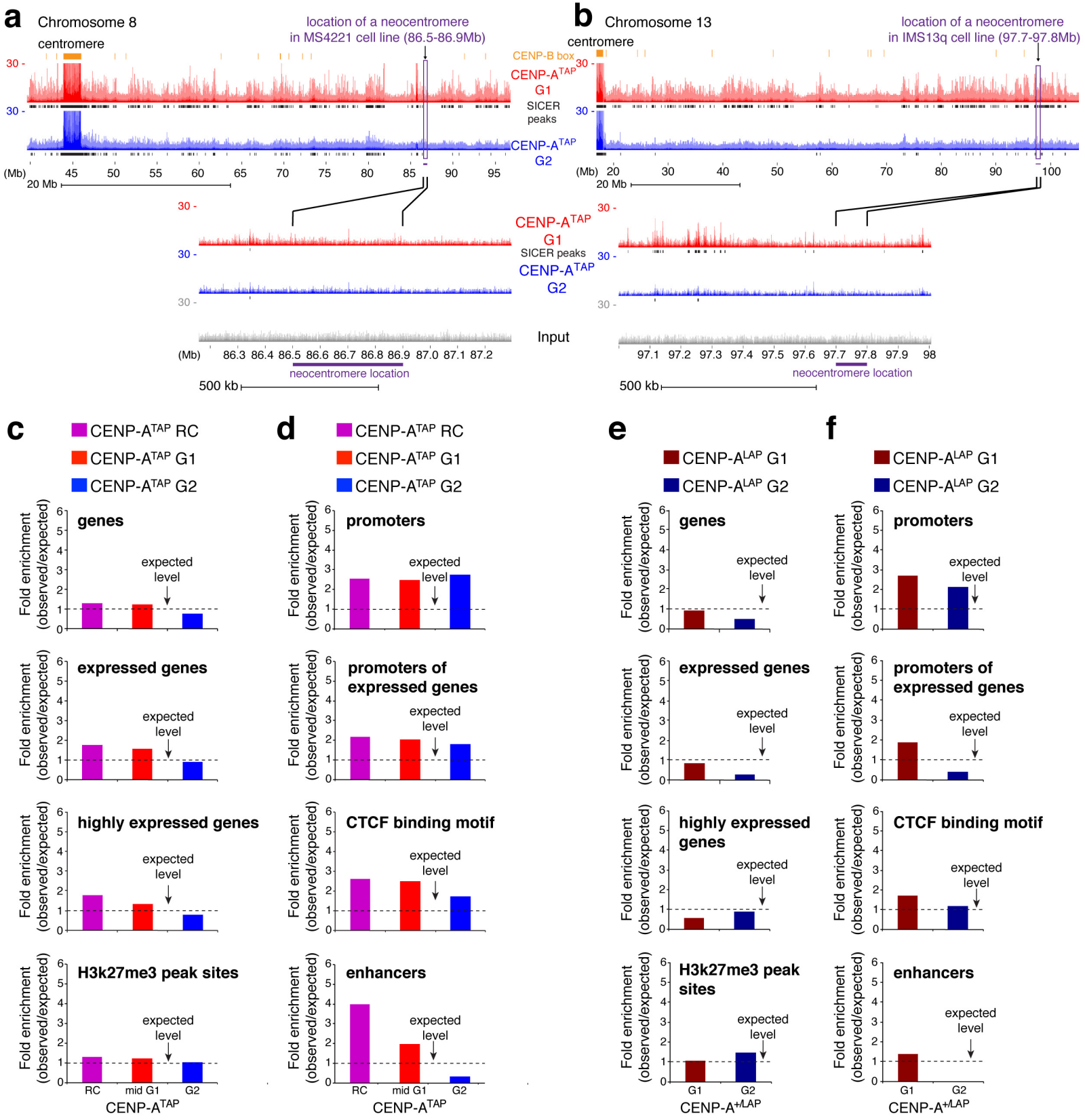


Fig S1. Identification of peaks enriched for CENP-A binding.

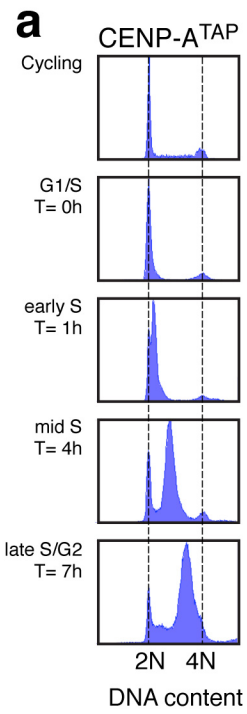
Nechemia-Arbely et al., Supplementary Figure S2

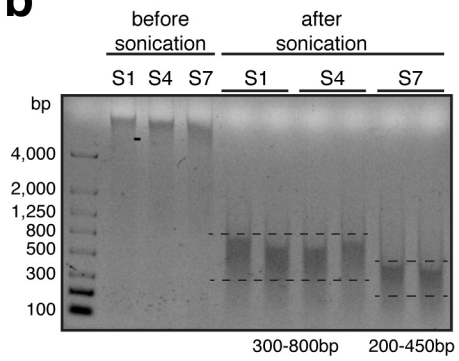


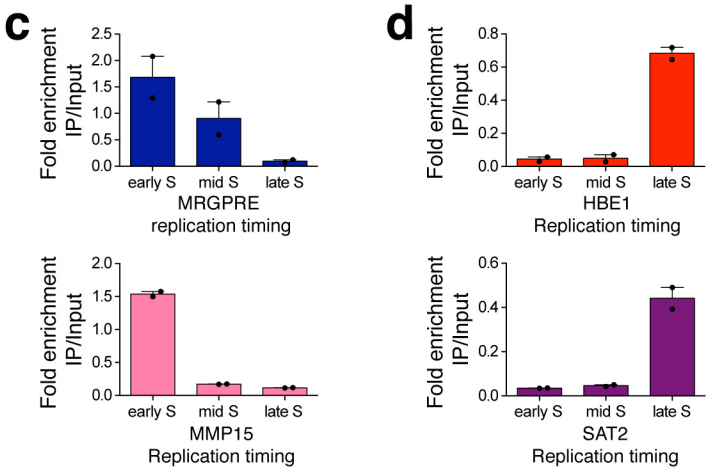
Supplementarry Figure S2. CENP-A ChIP-seq identifies CENP-A binding at reference centromeres of 23 human chromosomes

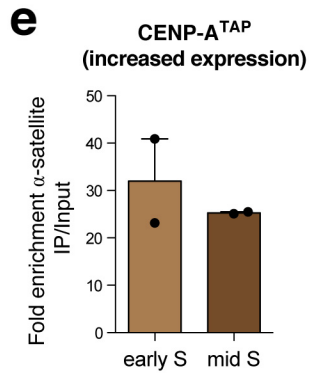


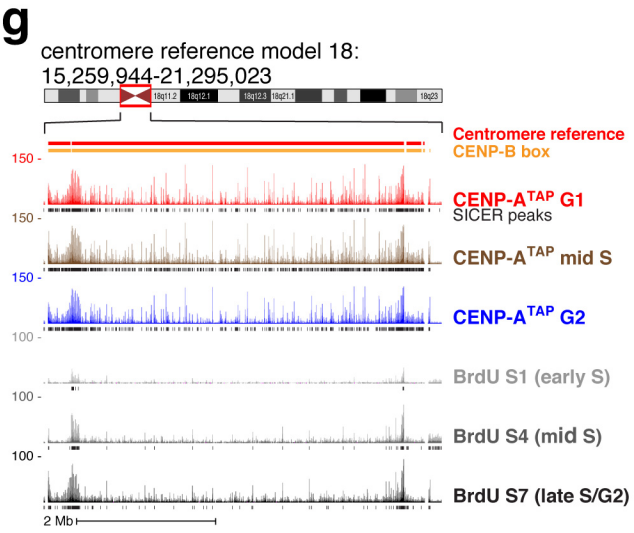
Supplementary Fig S3. Ectopic deposition of CENP-A into open and active chromatin at G1 does not function as a seeding hotspot for neocentromere formation

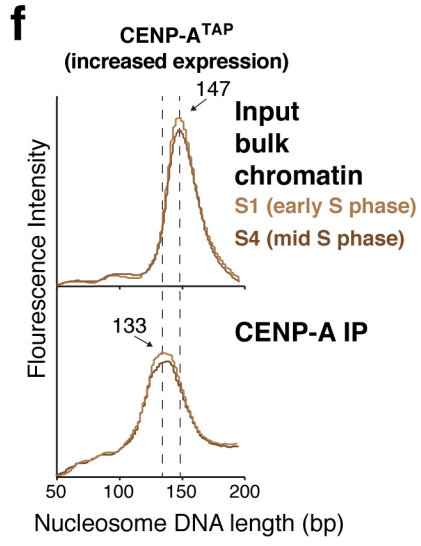


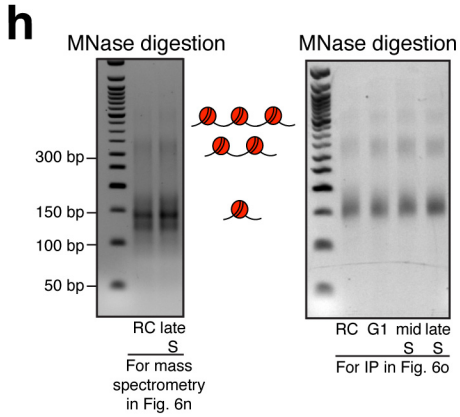












i MCM-2-7 and CAF1 complexes

	Randomly cycling spectra counts	Randomly cycling peptide counts	late S/G2 spectra counts	late S/G2 peptide counts
MCM2	0	0	15	11
MCM3	19	10	26	15
MCM4	27	11	31	23
MCM5	20	9	30	17
MCM6	17	8	16	8
MCM7	71	22	79	32
CAF1p150	0	0	2	2
CAF1p48	37	8	19	12

j DNA replication related proteins

	Randomly cycling spectra counts	Randomly cycling peptide counts	late S/G2 spectra counts	late S/G2 peptide counts
RFC1	85	30	31	14
RFC2	38	14	29	13
RFC3	24	13	14	9
RFC4	41	17	27	16
RFC5	34	11	21	11
GINS2	0	0	4	3
GINS3	2	2	3	3
POLA1	7	5	9	9
POLD1	4	4	17	12
ORC1	24	19	15	13
ORC5	5	2	12	6
PCNA	6	5	3	3

k Chromatin remodeling factors and nuclear chaperones

	Randomly cycling spectra counts	Randomly cycling peptide counts	late S/G2 spectra counts	late S/G2 peptide counts
HJURP	469	90	430	87
NPM1	415	45	282	43
RbAp46	29	13	19	11
NPM3	10	5	8	3
RSF1	306	79	73	39
FACTp80	430	79	181	59
FACTp140	1725	147	434	107
DAXX	12	7	0	0
NAP1L1	15	6	25	8
NAP1L4	4	3	5	3

l Histones

	Randomly cycling spectra counts	Randomly cycling peptide counts	late S/G2 spectra counts	late S/G2 peptide counts
H2A	1,924	96	810	81
H2B	520	31	205	25
H3.1	245	26	93	22
H4	737	48	396	66
H1	26	15	12	8

m centromere and kinetochore

	Randomly cycling spectra counts	Randomly cycling peptide counts	late S/G2 spectra counts	late S/G2 peptide counts
AURKB	8	7	16	12
Borealin	10	6	9	9
INCENP	12	10	14	13
KNL1	98	47	0	0
ZWINT	16	7	25	9
NDC80	5	3	48	26
NUF2	0	0	6	6
SPC24	0	0	19	7
SPC25	11	4	22	9
DSN1	47	18	47	19
MIS12	19	9	16	9
PMF1	20	11	19	11
NSL1	81	17	47	12
Mis18α	15	5	0	0
Mis18β	14	7	9	6
Mis18bp1	36	22	27	19

Supplementary Figure S4. Centromeres are late replicating with CENP-A remaining tethered locally by continued binding to the CCAN complex

Supplementary Table S1. Read statistics for ChIP-seq and Repli-seq experiments. Total number of merged paired-end read (one read per merged two paired-ends) generated for each sample in dataset, the number (and percentage) of those that were >=100bp in length and the number (and percentage) of reads mapping to α-satellites. **Table A, B.** Read statistics for each sample in the CENP-A$^{+/LAP}$ dataset (A) and for the combined replicates in the CENP-A$^{+/LAP}$ dataset (D) in each condition. **Table C, D.** Read statistics for each sample in the CENP-A^{TAP} dataset (C) and for the combined replicates in the CENP-A^{TAP} dataset (D) in each condition. **Table E, F.** Read statistics for each sample in the BrdU Repli-seq dataset (E) and for the combined replicates in the BrdU Repli-seq dataset (F) in each condition.

Table A. ChIP-seq replicates statistics for CENP-A$^{+/LAP}$ chromatin

Experiment	Replicate No	Total number of merged paired-end reads (100bp x 2)	Total (%) number of merged reads >=100bp	No (%) of merged reads mapping to α-satellites
CENP-A^{LAP} G1	1	37,088,538	30,232,099 (81.5%)	27,671,623 (74.6%)
	2	40,641,911	33,438,763 (82.3%)	30,342,295 (74.6%)
CENP-A^{LAP} G2	1	39,939,734	32,202,885 (80.6%)	24,209,577 (60.6%)
	2	32,689,317	25,933,735 (79.3%)	19,566,369 (59.9%)
CENP-A^{LAP} G1 Input	1	31,874,876	28,835,317 (90.5%)	952,554 (2.98%)

Table B. ChIP-seq combined replicate statistics for CENP-A$^{+/LAP}$ chromatin

Experiment	Total number of merged paired-end reads (100bp x 2)	Total (%) number of merged reads >=100bp	No (%) of merged reads mapping to α-satellites
CENP-A^{LAP} G1	77,730,449	63,670,862 (81.9%)	58,013,918 (74.6%)
CENP-A^{LAP} G2	72,629,051	58,136,620 (79.9%)	43,775,946 (60.2%)

Table C. ChIP-seq replicates statistics for CENP-A^{TAP} chromatin

Experiment	Replicate No	Total number of merged paired-end reads (100bp x 2)	Total (%) number of merged reads >=100bp	No (%) of merged reads mapping to α-satellites
CENP-A^{TAP} RC	1	9,436,346	8,464,391 (89.7%)	4,588,228 (48.6%)
	2	69,039,423	54,952,932 (79.6%)	31,053,433 (45%)
CENP-A^{TAP} G1	1	68,776,382	54,522,683 (79.3%)	28,023,214 (40.7%)
	2	51,746,426	40,328,176 (77.9%)	21,534,570 (41.6%)
CENP-A^{TAP} mid S	1	65,077,481	50,476,007 (77.6%)	33,879,664 (52.1%)
	2	62,298,174	48,772,959 (78.3%)	31,783,490 (51.0%)
CENP-A^{TAP} G2	1	49,206,313	40,088,665 (81.5%)	25,204,739 (51.2%)
	2	60,985,667	48,764,504 (80.0%)	33,414,704 (54.8%)
H3.1^{TAP} RC	1	31,819,170	27,823,087 (87.4%)	668,075 (2.1%)
	2	42,069,089	36,187,314 (86.0%)	911,027 (2.2%)
CENP-A^{TAP} RC Input	1	61,557,119	61,612,349 (84.5%)	1,374,644 (2.2%)

Table D. ChIP-seq combined replicate statistics for CENP-A^{TAP} chromatin

Experiment	Total number of merged paired-end reads (100bp x 2)	Total (%) number of merged reads >=100bp	No (%) of merged reads mapping to α-satellites
CENP-A^{TAP} RC	78,475,769	63,417,323 (80.8%)	35,641,661 (46.8%)
CENP-A^{TAP} G1	120,522,808	94,850,859 (78.7%)	49,557,784 (41.2%)
CENP-A^{TAP} mid S	127,375,655	99,248,966 (77.9%)	65,663,154 (51.5%)
CENP-A^{TAP} G2	110,191,980	88,853,169 (80.6%)	58,619,443 (53%)
H3.1^{TAP} RC	73,888,259	64,010,401 (86.6%)	1,579,102 (2.15%)

Table E. Repli-seq replicates statistics for BrdU labeled DNA

Experiment	Replicate No	Total number of merged paired-end reads (100bp x 2)	Total (%) number of merged reads >=100bp	No (%) of merged reads mapping to α-satellites
BrdU S1 Early S phase	1	48,838,225	42,840,706 (87.7%)	144,337 (0.29%)
	2	938,746	730,915 (77.8%)	2,082 (0.22%)
BrdU S4 Mid S phase	1	46,828,553	40,991,823 (87.5%)	371,793 (0.79%)
	2	34,480,560	31,507,686 (91.4%)	254,159 (0.73%)
BrdU S7 Late S phase	1	40,899,839	35,827,676 (87.6%)	1,657,902 (4.05%)
	2	41750,126	35,667,382 (85.4%)	1,682,353 (4.03%)
BrdU S1 Input	1	22,887,332	21,177,083 (92.5%)	1,185,850 (5.2%)
BrdU S4 Input	1	25,806,345	23,810,449 (92.2%)	1,146,797 (4.4%)
BrdU S7 Input	1	25,004,047	23,205,784 (92.8%)	1,322,554 (5.3%)

Table F. Repli-seq combined replicate statistics for BrdU labeled DNA

Experiment	Total number of merged paired-end reads (100bp x 2)	Total (%) number of merged reads >=100bp	No (%) of merged reads mapping to α-satellites
BrdU S1 Early S phase	49,776,971	43,571,621 (82.8%)	146,419 (0.25%)
BrdU S4 Mid S phase	81,309,113	72,499,509 (89.4%)	625,952 (0.76%)
BrdU S7 Late S phase	82,649,965	71,495,058 (86.5%)	3,340,255 (4.04%)

Table S2. Endogenous CENP-A sequence mapping onto α-satellite DNAs in human centromere reference models for each autosome and the X chromosome. Centromere reference models are from Miga et al. ([46], unpublished), generated with methods as previously described [45]. Length estimates are expected to be averaged across arrays from homologous chromosomes. **Column 1:** chromosome information, **column 2:** chromosome start position, **column 3:** chromosome end position, **column 4:** length in bp of each reference model as represented in the human assembly [46, 8], **column 5:** Genbank accession, **columns 6:** Genomic locus, if applicable, **column 7, 8, 9:** number of reads for CENP-A^{LAP} G1, replicate samples 1 and 2, and input, respectively, that aligned to the α-satellite reference model, **columns 10, 11, 12:** relative frequency of alignment to the α-satellite reference model is given for CENP-A^{LAP} G1, replicate samples 1 and 2, and input, respectively. **Columns 13, 14:** fold-enrichment of CENP-A^{LAP} G1, replicate samples 1 and 2 at the α-satellite reference model, relative to input. A summary of the reads and bases is given for those chromosomes that have several α-satellite reference models. Arrays that are identical between different chromosome locations are indicated as follows: *Sum of three near-identical arrays on chr1, 5, and 19; **Sum of two near-identical arrays on chr5, and 19; ***Sum of acrocentric near-identical arrays on chr13, 14, 21 and 22. Sequence coordinates refer to the human GRCh38 assembly.

hg38 chromo-some number	chromosome coordinates Start	chromosome coordinates End	Length (bp)	GenBank Accession number	Locus	CENP-A G1-1 Read count	CENP-A G1-2 Read count	Input Read count	CENP-A G1-1 Relative Frequency	CENP-A G1-2 Relative Frequency	Input Relative Frequency	CENP-A G1-1 Enrichment	CENP-A G1-2 Enrichment
chr1	122,026,459	122,224,535	198,076	GJ211836.1	N/A	465	400	354	1.29E-05	1.05E-05	1.11E-05	1.16	0.94
chr1	122,224,635	122,503,147	278,512	GJ211837.1	D1Z5	6,636	6,492	2,363	1.84E-04	1.70E-04	7.43E-05	2.48	2.29
chr1*	122,503,247	124,785,432	2,282,185	GJ212202.1	D1Z7/ D5Z2/ D19Z3	3,884,766	4,104,057	153,199	1.08E-01	1.08E-01	4.82E-03	**22.41**	**22.35**
chr1	124,785,532	124,849,129	63,597	GJ211855.1	N/A	1,154	1,189	797	3.21E-05	3.12E-05	2.51E-05	1.28	1.24
chr1	124,849,229	124,932,724	83,495	GJ211857.1	N/A	4,408	4,565	965	1.22E-04	1.20E-04	3.03E-05	4.04	3.95
Sum chr1			**2,905,865**			**3,897,429**	**4,116,703**	**157,678**	**1.08E-01**	**1.08E-01**	**4.96E-03**	**21.84**	**21.78**
chr2	**92,188,145**	**94,090,557**	**1,902,412**	**GJ211860.1**	**D2Z1**	**772,858**	**773,709**	**13,622**	**2.15E-02**	**2.03E-02**	**4.28E-04**	**50.14**	**47.39**
chr3	90,772,458	91,233,586	461,128	GJ211866.1	N/A	3,287	2,626	3,396	9.13E-05	6.89E-05	1.07E-04	0.86	0.65
chr3	91,233,686	91,247,622	13,936	GJ211867.1	N/A	90	93	146	2.50E-06	2.44E-06	4.59E-06	0.54	0.53
chr3	91,553,419	93,655,574	2,102,155	GJ211871.1	D3Z1	726,106	743,093	59,085	2.02E-02	1.95E-02	1.86E-03	**10.86**	**10.49**
Sum chr3			**2,577,219**			**729,483**	**745,812**	**62,627**	**2.03E-02**	**1.96E-02**	**1.97E-03**	**10.29**	**9.94**
chr4	**49,712,061**	**51,743,951**	**2,031,890**	**GJ211881.1**	**D4Z1**	**802,017**	**880,598**	**25,601**	**2.23E-02**	**2.31E-02**	**8.05E-04**	**27.69**	**28.70**
chr5	46,485,900	46,569,062	83,162	GJ211882.1	N/A	1,138	1,262	727	3.16E-05	3.31E-05	2.29E-05	1.38	1.45
chr5	46,569,162	46,796,725	227,563	GJ211883.1	N/A	1,055	1,064	632	2.93E-05	2.79E-05	1.99E-05	1.48	1.40
chr5	46,796,825	47,061,288	264,463	GJ211884.1	N/A	2,175	2,406	1,712	6.04E-05	6.31E-05	5.38E-05	1.12	1.17
chr5	47,106,994	47,153,339	46,345	GJ211886.1	N/A	1,180	1,321	864	3.28E-05	3.47E-05	2.72E-05	1.21	1.28
chr5	47,153,439	47,296,069	142,630	GJ211887.1	N/A	1,589	1,669	361	4.42E-05	4.38E-05	1.14E-05	3.89	3.86

chr5*	47,309,184	49,591,369	2,282,185	GJ212203.1	D1Z7/ D5Z2/ D19Z3	3,884,766	4,104,057	153,199	1.08E-01	1.08E-01	4.82E-03	**22.41**	**22.35**
chr5**	49,667,531	49,721,203	53,672	GJ211904.2	N/A	421	377	464	1.17E-05	9.89E-06	1.46E-05	0.80	0.68
chr5	49,721,303	50,059,807	338,504	GJ211906.2	N/A	1,094	1,196	1,456	3.04E-05	3.14E-05	4.58E-05	0.66	0.69
Sum chr5			**3,438,524**			**3,893,418**	**4,113,352**	**159,415**	**1.08E-01**	**1.08E-01**	**5.01E-03**	**21.58**	**21.53**
chr6	**58,553,888**	**59,829,934**	**1,276,046**	**GJ211907.1**	**D6Z1**	**1,240,000**	**1,360,728**	**41,164**	**3.45E-02**	**3.57E-02**	**1.29E-03**	**26.62**	**27.58**
chr7	58,169,653	60,828,234	2,658,581	GJ211908.1	D7Z1	987,890	1,080,834	25,905	2.75E-02	2.84E-02	8.15E-04	**33.70**	**34.81**
chr7	61,377,788	61,528,020	150,232	GJ212194.1	D7Z2	1,818	1,897	890	5.05E-05	4.98E-05	2.80E-05	1.81	1.78
Sum chr7			**2,808,813**			**989,708**	**1,082,731**	**26,795**	**2.75E-02**	**2.84E-02**	**8.43E-04**	**32.64**	**33.71**
chr8	**44,033,744**	**45,877,265**	**1,843,521**	**GJ211909.1**	**D8Z2**	**962,237**	**1,055,322**	**21,112**	**2.67E-02**	**2.77E-02**	**6.64E-04**	**40.28**	**41.70**
chr9	**43,389,635**	**45,518,558**	**2,128,923**	**GJ211929.1**	**D9Z4**	**726,448**	**801,194**	**16,178**	**2.02E-02**	**2.10E-02**	**5.09E-04**	**39.68**	**41.32**
chr10	39,686,682	39,935,900	249,218	GJ211930.1	N/A	215,957	211,519	859	6.00E-03	5.55E-03	2.70E-05	**222.18**	**205.43**
chr10	39,936,000	41,497,440	1,561,440	GJ211932.1	D10Z1	503,166	508,384	14,260	1.40E-02	1.33E-02	4.48E-04	31.18	29.74
chr10	41,497,540	41,545,720	48,180	GJ211933.1	N/A	18,519	16,696	1,314	5.15E-04	4.38E-04	4.13E-05	12.46	10.60
chr10	41,545,820	41,593,521	47,701	GJ211936.1	N/A	4,002	3,829	1,465	1.11E-04	1.00E-04	4.61E-05	2.41	2.18
Sum chr10			**1,906,539**			**741,644**	**740,428**	**17,898**	**2.06E-02**	**1.94E-02**	**5.63E-04**	**36.62**	**34.51**
chr11	51,078,348	51,090,317	11,969	GJ211938.1	N/A	63	23	29	1.75E-06	6.03E-07	9.12E-07	1.92	0.66
chr11	51,090,417	54,342,399	3,251,982	GJ211943.1	D11Z1	811,562	827,321	35,762	2.26E-02	2.17E-02	1.12E-03	**20.06**	**19.30**
chr11	54,342,499	54,425,074	82,575	GJ211948.1	N/A	74,089	75,567	2,855	2.06E-03	1.98E-03	8.98E-05	**22.93**	**22.08**
Sum chr11			**3,346,526**			**885,714**	**902,911**	**38,646**	**2.46E-02**	**2.37E-02**	**1.22E-03**	**20.25**	**19.49**
chr12	34,769,407	34,816,611	47,204	GJ211949.1	N/A	773	477	274	2.15E-05	1.25E-05	8.62E-06	2.49	1.45
chr12	34,835,295	37,185,252	2,349,957	GJ211954.1	D12Z3	1,559,732	1,576,847	59,791	4.33E-02	4.14E-02	1.88E-03	**23.05**	**22.00**
Sum chr12			**2,397,161**			**1,560,505**	**1,577,324**	**60,065**	**4.34E-02**	**4.14E-02**	**1.89E-03**	**22.96**	**21.91**
chr13***	16,000,000	16,022,537	22,537	GJ211955.2	N/A	417	417	429	1.16E-05	1.09E-05	1.35E-05	0.86	0.81
chr13***	16,022,637	16,110,659	88,022	GJ211961.2	N/A	556	529	587	1.54E-05	1.39E-05	1.85E-05	0.84	0.75
chr13***	16,110,759	16,164,892	54,133	GJ211962.2	N/A	2,684	2,706	3,491	7.46E-05	7.10E-05	1.10E-04	0.68	0.65
chr13***	16,164,992	16,228,527	63,535	GJ211963.2	N/A	1,392	1,369	1,549	3.87E-05	3.59E-05	4.87E-05	0.79	0.74
chr13***	16,228,627	16,249,297	20,670	GJ211965.2	N/A	1,621	1,627	1,914	4.50E-05	4.27E-05	6.02E-05	0.75	0.71
chr13***	16,249,397	16,256,067	6,670	GJ211967.2	N/A	2,773	2,699	1,877	7.71E-05	7.08E-05	5.90E-05	1.31	1.20

chr13***	16,256,167	16,259,412	3,245	GJ211968.2	N/A	64	61	53	1.78E-06	1.60E-06	1.67E-06	1.07	0.96
chr13***	16,259,512	16,282,073	22,561	GJ211969.2	N/A	290	294	274	8.06E-06	7.71E-06	8.62E-06	0.94	0.90
chr13***	16,282,173	17,416,384	1,134,211	GJ211972.2	N/A	2,485,354	2,490,600	50,193	6.91E-02	6.53E-02	1.58E-03	**43.76**	**41.40**
chr13***	17,416,484	17,416,824	340	GJ212205.1	N/A	12	15	1	3.33E-07	3.93E-07	3.14E-08	**10.60**	**12.51**
chr13***	17,416,924	17,417,264	340	GJ212206.1	N/A	29	34	1	8.06E-07	8.92E-07	3.14E-08	**25.63**	**28.37**
chr13***	17,417,364	17,418,562	1,198	GJ211986.2	N/A	576	563	432	1.60E-05	1.48E-05	1.36E-05	1.18	1.09
chr13***	17,418,662	18,051,248	632,586	GJ211991.2	D13Z1/ D21Z1 alphaRI, L1.26	1,535,634	1,543,548	24,960	4.27E-02	4.05E-02	7.85E-04	**54.37**	**51.59**
Sum chr13			**2,050,048**			**4,031,402**	**4,044,462**	**85,761**	**1.12E-01**	**1.06E-01**	**2.70E-03**	**41.54**	**39.34**
chr14***	16,000,000	16,022,537	22,537	GJ211992.2	N/A	417	417	429	1.16E-05	1.09E-05	1.35E-05	0.86	0.81
chr14***	16,140,627	16,228,649	88,022	GJ211998.2	N/A	556	529	587	1.54E-05	1.39E-05	1.85E-05	0.84	0.75
chr14***	16,228,749	16,282,882	54,133	GJ211999.2	N/A	2,684	2,706	3,491	7.46E-05	7.10E-05	1.10E-04	0.68	0.65
chr14***	16,282,982	16,346,517	63,535	GJ212000.2	N/A	1,392	1,369	1,549	3.87E-05	3.59E-05	4.87E-05	0.79	0.74
chr14***	16,346,617	16,367,287	20,670	GJ212002.2	N/A	1,621	1,627	1,914	4.50E-05	4.27E-05	6.02E-05	0.75	0.71
chr14***	16,367,387	16,374,057	6,670	GJ212004.2	N/A	2,773	2,699	1,877	7.71E-05	7.08E-05	5.90E-05	1.31	1.20
chr14***	16,374,157	16,377,402	3,245	GJ212005.2	N/A	64	61	53	1.78E-06	1.60E-06	1.67E-06	1.07	0.96
chr14***	16,377,502	16,400,063	22,561	GJ212006.2	N/A	290	294	274	8.06E-06	7.71E-06	8.62E-06	0.94	0.90
chr14***	16,404,448	17,538,659	1,134,211	GJ212009.2	N/A	2,485,354	2,490,600	50,193	6.91E-02	6.53E-02	1.58E-03	**43.76**	**41.40**
chr14***	17,538,759	17,539,099	340	GJ212210.1	N/A	12	15	1	3.33E-07	3.93E-07	3.14E-08	**10.60**	**12.51**
chr14***	17,539,199	17,539,539	340	GJ212211.1	N/A	29	34	1	8.06E-07	8.92E-07	3.14E-08	**25.63**	**28.37**
chr14***	17,539,639	17,540,837	1,198	GJ212023.2	N/A	576	563	432	1.60E-05	1.48E-05	1.36E-05	1.18	1.09
chr14***	17,540,937	18,173,523	632,586	GJ212028.2	N/A	1,535,634	1,543,548	24,960	4.27E-02	4.05E-02	7.85E-04	**54.37**	**51.59**
Sum chr14			**2,050,048**			**4,031,402**	**4,044,462**	**85,761**	**1.12E-01**	**1.06E-01**	**2.70E-03**	**41.54**	**39.34**
chr15	17,083,673	17,498,951	415,278	GJ212036.1	N/A	1,740	1,718	1,281	4.83E-05	4.51E-05	4.03E-05	1.20	1.12
chr15	17,499,051	18,355,008	855,957	GJ212042.1	N/A	9,617	9,910	4,554	2.67E-04	2.60E-04	1.43E-04	1.87	1.82
chr15	18,355,108	19,725,254	1,370,146	GJ212045.1	D15Z3	1,195,175	1,187,663	14,547	3.32E-02	3.12E-02	4.57E-04	**72.61**	**68.11**
Sum chr15			**2,641,381**			**1,206,532**	**1,199,291**	**20,382**	**3.35E-02**	**3.15E-02**	**6.41E-04**	**52.31**	**49.09**
chr16	36,311,158	36,334,460	23,302	GJ212046.1	N/A	419	387	147	1.16E-05	1.02E-05	4.62E-06	2.52	2.20

chr16	36,337,666	38,265,669	1,928,003	GJ212051.1	D16Z2	857,505	872,514	19,641	2.38E-02	2.29E-02	6.18E-04	**38.58**	**37.06**
Sum chr16			**1,951,305**			**857,924**	**872,901**	**19,788**	**2.38E-02**	**2.29E-02**	**6.22E-04**	**38.32**	**36.80**
chr17	22,813,679	23,194,918	381,239	GJ212053.1	D17Z1B	48,160	47,722	2,792	1.34E-03	1.25E-03	8.78E-05	**15.24**	**14.26**
chr17	23,195,018	26,566,633	3,371,615	GJ212054.1	D17Z1	358,116	338,203	11,231	9.95E-03	8.87E-03	3.53E-04	**28.18**	**25.12**
chr17	26,566,733	26,616,164	49,431	GJ212055.1	N/A	87,241	90,634	2,765	2.42E-03	2.38E-03	8.69E-05	**27.88**	**27.35**
Sum chr17			**3,802,285**			**493,517**	**476,559**	**16,788**	**1.37E-02**	**1.25E-02**	**5.28E-04**	**25.98**	**23.68**
chr18	15,460,899	15,780,377	319,478	GJ212060.1	N/A	16,312	15,458	2,974	4.53E-04	4.05E-04	9.35E-05	4.85	4.34
chr18	15,797,855	20,561,439	4,763,584	GJ212062.1	D18Z1	530,474	534,958	16,477	1.47E-02	1.40E-02	5.18E-04	**28.45**	**27.09**
chr18	20,603,247	20,696,289	93,042	GJ212066.1	D18Z2	2,198	2,120	737	6.11E-05	5.56E-05	2.32E-05	2.64	2.40
chr18	20,696,389	20,736,025	39,636	GJ212067.1	N/A	503	386	127	1.40E-05	1.01E-05	3.99E-06	3.50	2.54
chr18	20,736,125	20,813,083	76,958	GJ212069.1	N/A	812	650	170	2.26E-05	1.71E-05	5.35E-06	4.22	3.19
chr18	20,839,797	20,861,206	21,409	GJ212071.1	N/A	120	92	32	3.33E-06	2.41E-06	1.01E-06	3.31	2.40
Sum chr18			**5,314,107**			**550,419**	**553,664**	**20,517**	**1.53E-02**	**1.45E-02**	**6.45E-04**	**23.71**	**22.51**
chr19**	24,498,980	24,552,652	53,672	GJ212072.2	N/A	421	377	464	1.17E-05	9.89E-06	1.46E-05	0.80	0.68
chr19	24,552,752	24,891,256	338,504	GJ212077.2	N/A	1,176	1,165	1,617	3.27E-05	3.06E-05	5.08E-05	0.64	0.60
chr19*	24,908,689	27,190,874	2,282,185	GJ212201.1	D1Z7/ D5Z2/ D19Z3	3,884,766	4,104,057	153,199	1.08E-01	1.08E-01	4.82E-03	**22.41**	**22.35**
Sum chr19			**2,674,361**			**3,886,363**	**4,105,599**	**155,280**	**1.08E-01**	**1.08E-01**	**4.88E-03**	**22.12**	**22.06**
chr20	26,436,232	26,586,955	150,723	GJ212091.1	N/A	8,724	6,977	2,142	2.42E-04	1.83E-04	6.73E-05	3.60	2.72
chr20	26,608,145	28,494,539	1,886,394	GJ212093.1	D20Z2	578,724	588,963	11,545	1.61E-02	1.54E-02	3.63E-04	**44.30**	**42.56**
chr20	28,508,997	28,556,953	47,956	GJ212095.1	N/A	3,688	3,180	2,219	1.02E-04	8.34E-05	6.98E-05	1.47	1.20
chr20	28,648,108	28,728,874	80,766	GJ212105.1	N/A	279	276	378	7.75E-06	7.24E-06	1.19E-05	0.65	0.61
chr20	29,125,793	29,204,668	78,875	GJ212107.1	N/A	174	136	204	4.83E-06	3.57E-06	6.41E-06	0.75	0.56
chr20	29,917,404	30,038,348	120,944	GJ212117.1	N/A	379	246	297	1.05E-05	6.45E-06	9.34E-06	1.13	0.69
Sum chr20			**2,365,658**			**591,968**	**599,778**	**16,785**	**1.64E-02**	**1.57E-02**	**5.28E-04**	**31.17**	**29.81**
chr21***	10,864,560	10,887,097	22,537	GJ212118.2	N/A	417	417	429	1.16E-05	1.09E-05	1.35E-05	0.86	0.81
chr21***	10,887,197	10,975,219	88,022	GJ212124.2	N/A	556	529	587	1.54E-05	1.39E-05	1.85E-05	0.84	0.75
chr21***	10,975,319	11,029,452	54,133	GJ212125.2	N/A	2,684	2,706	3,491	7.46E-05	7.10E-05	1.10E-04	0.68	0.65
chr21***	11,029,552	11,093,087	63,535	GJ212126.2	N/A	1,392	1,369	1,549	3.87E-05	3.59E-05	4.87E-05	0.79	0.74

chr21***	11,093,187	11,113,857	20,670	GJ212128.2	N/A	1,621	1,627	1,914	4.50E-05	4.27E-05	6.02E-05	0.75	0.71
chr21***	11,113,957	11,120,627	6,670	GJ212130.2	N/A	2,773	2,699	1,877	7.71E-05	7.08E-05	5.90E-05	1.31	1.20
chr21***	11,120,727	11,123,972	3,245	GJ212131.2	N/A	64	61	53	1.78E-06	1.60E-06	1.67E-06	1.07	0.96
chr21***	11,124,072	11,146,633	22,561	GJ212132.2	N/A	290	294	274	8.06E-06	7.71E-06	8.62E-06	0.94	0.90
chr21***	11,146,733	12,280,944	1,134,211	GJ212135.2	N/A	2,485,354	2,490,600	50,193	6.91E-02	6.53E-02	1.58E-03	**43.76**	**41.40**
chr21***	12,281,044	12,281,384	340	GJ212204.1	N/A	12	15	1	3.33E-07	3.93E-07	3.14E-08	**10.60**	**12.51**
chr21***	12,281,484	12,281,824	340	GJ212207.1	N/A	29	34	1	8.06E-07	8.92E-07	3.14E-08	**25.63**	**28.37**
chr21***	12,281,924	12,283,122	1,198	GJ212149.2	N/A	576	563	432	1.60E-05	1.48E-05	1.36E-05	1.18	1.09
chr21***	12,283,222	12,915,808	632,586	GJ212154.2	D13Z1/ D21Z1 alphaRI, L1.26	1,535,634	1,543,548	24,960	4.27E-02	4.05E-02	7.85E-04	**54.37**	**51.59**
Sum chr21			**2,050,048**			**4,031,402**	**4,044,462**	**85,761**	**1.12E-01**	**1.06E-01**	**2.70E-03**	**41.54**	**39.34**
chr22***	12,954,788	12,977,325	22,537	GJ212155.2	N/A	417	417	429	1.16E-05	1.09E-05	1.35E-05	0.86	0.81
chr22***	13,021,422	13,109,444	88,022	GJ212161.2	N/A	556	529	587	1.54E-05	1.39E-05	1.85E-05	0.84	0.75
chr22***	13,109,544	13,163,677	54,133	GJ212162.2	N/A	2,684	2,706	3,491	7.46E-05	7.10E-05	1.10E-04	0.68	0.65
chr22***	13,163,777	13,227,312	63,535	GJ212163.2	N/A	1,392	1,369	1,549	3.87E-05	3.59E-05	4.87E-05	0.79	0.74
chr22***	13,227,412	13,248,082	20,670	GJ212165.2	N/A	1,621	1,627	1,914	4.50E-05	4.27E-05	6.02E-05	0.75	0.71
chr22***	13,248,182	13,254,852	6,670	GJ212167.2	N/A	2,773	2,699	1,877	7.71E-05	7.08E-05	5.90E-05	1.31	1.20
chr22***	13,254,952	13,258,197	3,245	GJ212168.2	N/A	64	61	53	1.78E-06	1.60E-06	1.67E-06	1.07	0.96
chr22***	13,258,297	13,280,858	22,561	GJ212169.2	N/A	290	294	274	8.06E-06	7.71E-06	8.62E-06	0.94	0.90
chr22***	13,285,243	14,419,454	1,134,211	GJ212172.2	N/A	2,485,354	2,490,600	50,193	6.91E-02	6.53E-02	1.58E-03	**43.76**	**41.40**
chr22***	14,419,554	14,419,894	340	GJ212208.1	N/A	12	15	1	3.33E-07	3.93E-07	3.14E-08	**10.60**	**12.51**
chr22***	14,419,994	14,420,334	340	GJ212209.1	N/A	29	34	1	8.06E-07	8.92E-07	3.14E-08	**25.63**	**28.37**
chr22***	14,420,434	14,421,632	1,198	GJ212186.2	N/A	576	563	432	1.60E-05	1.48E-05	1.36E-05	1.18	1.09
chr22***	14,421,732	15,054,318	632,586	GJ212191.2	N/A	1,535,634	1,543,548	24,960	4.27E-02	4.05E-02	7.85E-04	**54.37**	**51.59**
Sum chr22			**2,050,048**			**4,031,402**	**4,044,462**	**85,761**	**1.12E-01**	**1.06E-01**	**2.70E-03**	**41.54**	**39.34**
chrX	**58,605,579**	**62,412,542**	**3,806,963**	**GJ212192.1**	**DXZ1**	**725,621**	**794,981**	**17,185**	**2.02E-02**	**2.09E-02**	**5.40E-04**	**37.32**	**38.59**

Primary antibodies

Antibody	Species	Assay and dilution	Source	Catalogue	Validation
BrdU	Mouse	Repli-seq	Becton-Dickinson Biosciences	CAT # 555627	used for Repli-seq previously (Hansen et al., 2010)
GFP	Mouse	CENP-ALAP ChIP-seq	MSKCC antibody core facility	clone 19C8 and clone 19F7	validated previously by us and others
rabbit IgG	rabbit	CENP-ATAP ChIP-seq	Sigma-Aldrich	CAT # I5006	validated previously by us (Foltz et al., 2006) and others
CENP-A	rabbit	Immunoblot 1:1,000	Cell Signaling Technology	CAT # 2186	validated in Fig. 6o, also previously used by us and others
CAF1p150	rabbit	Immunoblot 1:500	Santa Cruz	CAT # sc-10772	validated in Fig. 6o; also see manufacturer's data sheet
CAF1p60	rabbit	Immunoblot 1:1,000	Bethyl Laboratories	CAT # A301-085A	validated in Fig. 6o; also see manufacturer's data sheet
CAF1p48	rabbit	Immunoblot 1:1,000	Bethyl Laboratories	CAT # A301-206A	validated in Fig. 6o; also see manufacturer's data sheet
MCM2	rabbit	Immunoblot 1:1,000	Abcam	CAT # Ab4461	validated in Fig. 6o; also see manufacturer's data sheet
CENP-B	rabbit	Immunofluorescence 1:1,000	Abcam	CAT # 25734	validated in Fig. S1d and previously by us and others
GFP	Mouse	Immunofluorescence 1:500	Roche	CAT # 11814460001	validated in Fig. S1c and previously by us and others
anti-centromere antibodies (ACA)	human	Immunofluorescence 1:500	Antibodies Inc	CAT # 15-234-0001	validated in Fig. S1c and previously by us and others

Secondary antibodies

Antibody	Species	Dilution	Source	Catalogue number
Sheep anti-mouse HRP	sheep	Immunoblot 1:4,000	GE Healthcare	NA931V
donkey anti-rabbit HRP	donkey	Immunoblot 1:4,000	GE Healthcare	NA934V
donkey anti-human TR	donkey	Immunofluorescence 1:300	Jackson Laboratories	CAT # 709-075-149
donkey anti-mouse FITC	donkey	Immunofluorescence 1:250	Jackson Laboratories	CAT # 715-095-151
FITC-rabbit IgG	rabbit	Immunofluorescence 1:200	Jackson Laboratories	CAT # 011-090-003